IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 11, Issue 4, April 2023

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

9940 572 462 | 6381 907 438 | ijircce@gmail.com | www.ijircce.com

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 4, April 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1104048 |

A Machine Learning Approach to Detecting and Preventing Cyberbullying Incidents

Sowjanya RN[*1], Dr. Sujatha SR[*2]

#II nd Year, Master of Technology, Department of Computer Science Engineering, Sri Siddhartha Institute of Technology, Tumakuru, Karnataka, India

*Professor, Department of Computer Science Engineering, Sri Siddhartha Institute of Technology, Tumakuru, Karnataka, India

ABSTRACT *:* The growing usage of digital and social media as a result of the development of technology has made communication as simple as clicking a button. Nevertheless, there is a bad side to the misuse of digital media in addition to the potential it presents. Today's culture is very concerned about cyberbullying, which is the act of spreading hatred and anger towards others through social media platforms. Cyberbullying can have devastating effects on a person's mental and emotional well-being, endangering their general wellbeing.

Despite the inevitable occurrence of cyberbullying, protective precautions can be taken. While existing solutions concentrate on identifying cyberbullying, they are difficult for the general public to use and do not take language usage changes into consideration, which might hinder detection. In order to close this gap, this research suggests a solution that uses supervised machine learning and takes important aspects of cyberbullying into account, such as the intent to injure, abusive language that is used repeatedly, and hate speech, for automatic identification and prevention.The technology recognises cyberbullying text and pinpoints related themes or groups, including racial or sexual insensitivity, physical cruelty, profanity, and other forms of abuse. Support vector machines and logistic regression are used for detection, and accuracy is improved by the use of well-established feature extraction methods as Term Frequency-Inverse Document Frequency (TF-IDF), N-gram, profanity analysis, and sentiment analysis. Recall, Precision, and F1-score measures are used to assess the proposed system's effectiveness in identifying cyberbullying.

KEYWORDS: Cyber Bullying, Term Frequency-Inverse Document Frequency, support vector machines, F1-scoremetrics, logistic regression, N-gram.

I. INTRODUCTION

The phrase "cyberbullying" didn't exist five years ago. Social media underwent a tremendous transition with the development of the World Wide Web, becoming quick and convenient to access on digital devices. The study by Hinduja and Patchin found that 36.5% of students have at least once been the victim of cyberbullying. Online comments that are cruel or hurtful are the most typical type of cyberbullying. In the USA, a poll of 1,501 kids aged 10 to 17 found thFacebook was shown to be the site where 80% of cyberbullying events involving Sri Lankans occurred. Unsettlingly, 65% of college students said that humiliating films or images of them had been posted, 15% admitted to uploading personal information, and 9% acknowledged disseminating fake material or lies.at 12% acknowledged to being abusive online, 4% reported being the victim of such abuse, and 3% reported being both bullies and victims. Around 1,000 instances of cyberbullying were reported in Sri Lanka by the Cyber Crimes Division (CID), with almost 90% of university students being exposed to it.Facebook was identified as the platform where Sri Lankans experienced 80% of cyberbullying incidents. Disturbingly, 65% of university students reported that embarrassing videos or photos of them were posted, 15% disclosed posting private information, and 9% admitted to spreading false information or lies.

The main problem with cyberbullying is how quickly it spreads to a large audience and how persistently it is seen over time. Schneider et alstudy, .'s which discovered a connection between cyberbullying victimisation and depressive symptoms (34%) and self-injury (24%), shows that this form of harassment has serious effects on victims' mental and physical health. But many cyberbullying cases go unreported, making it difficult for authorities to take action.In order to prevent cyberbullying, social media sites like Facebook, Instagram, and Twitter have put in place reporting and filtering tools as well as AI-powered hubs that ask users to think twice before publishing offensive content. Yet, these techniques can be difficult to monitor and are not always effective.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 4, April 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1104048 |

II. LITERATURE REVIEW

[1]. A supervised learning strategy was utilised by Yin et al.[1] to identify abuse from three different social media platforms. used the content features, sentiment features, and contextual features of documents coupled with the datasets from Myspace (discussion style communities), Slashdot, and Kongregate (chat-type communities). They employed a linear kernel lib SVM as a classification method. TFIDF weighting performed better than n-gram and foul language.
[2]. Three profanity detection algorithms were developed using a profanity dictionary, Levenshtein Edit Distance, and Bag-of-words [2]. The user-compiled list on phorum.com and noswearing.com served as the foundation for the profanity dictionary. The second technique employed a list of curse words in addition to an edit distance calculator to correct misspellings.To minimise false positives, the algorithm compares the words to a list of names and the English lexicon. The computer will categorise the word "shirt" as not profane by consulting the dictionary, despite the edit distance calculator matching the word "shirt" with the offending phrase "crap" and labelling it as such.
[3]. Bullies on Myspace and Spring were identified by Squicciarini et al. [3] using personal, social networking, and content-specific variables along with a Decision Tree classifier. We developed a set of guidelines to determine if a user's cyberbullying behaviour was sparked by the conduct of another bully.
[4]. A score was also created by Chavan and Shylaja [4] to indicate to other users the likelihood that a statement could offend them. They used a number of characteristics, such as skip-grams, and integrated the results of Support Vector Machine and Logistic Regression classifiers to enhance accuracy by 4% using a dataset from Kaggle10.

III. OBJECTIVES

- Retrieving current data from social media platforms.
- choose between Term Frequency-Inverse Frequency (TF-IF) and Count Vectorization (CV) as the optimal feature extraction approach for text categorization (TF-IDF)Building various Machine Learning Model using Training dataset.
- Deciding on the optimal model by weighing recall, f1 score, accuracy, and precision.
- make predictions based on test data and evaluate the outcomes.
- Development and implementation of a group chat application.

IV. EXISTING SYSTEM

The current state of cyberbullying detection systems does not effectively account for ironic content as a type of cyberbullying and lacks direct end-user control. Sarcastic content is not included in the definition of cyberbullying in the proposed approach, which intends to automatically identify and stop episodes of cyberbullying. In order to give a holistic strategy for addressing this issue, the system also identifies themes or categories linked to cyberbullying.

V. PROPOSED SYSTEM

Although machine learning algorithms might not be able to handle all features efficiently on their own, the suggested technique makes use of a combination of content-based features and sentiment-based characteristics. Whereas sentiment-based features pick up on the text's emotional undertone, content-based features are generated from the comments' actual content. This integrated strategy guarantees a thorough and thorough study of cyberbullying episodes for precise detection and prevention.

VI. PROPOSED SOLUTION

Recognizing that not all features may be amenable to machine learning techniques, the suggested method combines content-based features and sentiment-based data. Sentiment-based Features record the feelings the text conveys, while Content-based Features are generated from the literal content of the remarks. With this integrated strategy, occurrences of cyberbullying are thoroughly analysed, allowing for precise detection and prevention.
One of the Content-based Characteristics selected for the suggested strategy is TF-IDF (Term Frequency-Inverse Document Frequency). By taking into account both the term frequency (TF) and the inverse document frequency (IDF), the TF-IDF technique evaluates the importance of terms in a document (IDF). It considers a term's frequency in a document (TF) and scales it by the opposite of that term's frequency across all texts (IDF). This enables TF-IDF to ignore common terms like "is" and "am" that have little bearing on the overall results in favour of less frequent words that are more indicative of the substance and meaning of a document.

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

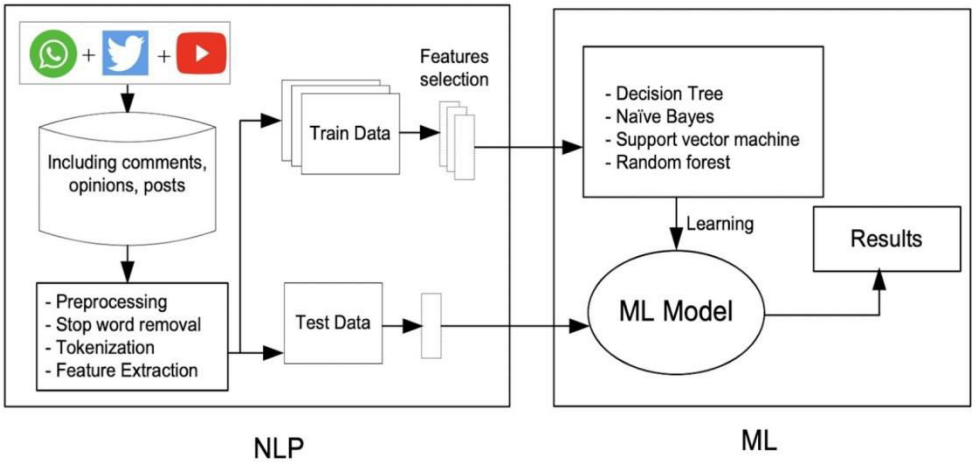
|| Volume 11, Issue 4, April 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1104048 |

VII. CYBERBULLYING DETECTION ARCHITECTURE

The abstract framework that describes the configuration and operation of a system is referred to as the system architecture. It provides an organised overview of the system's structural properties and acts as a blueprint or road map for it.An architecture description outlines the constituent parts or components of the system and offers instructions for locating the required goods and creating integrated systems that function well together to implement the system as a whole. It gives a clear and thorough overview of the structure of the system and how its parts work together to produce the required functionality and performance.

Fig 1: CyberBullying Detection Architecture



VIII. METHODOLOGY

The MVC (Model-View-Controller) design pattern was used to create the system architecture. Swing, a well-known Java GUI toolkit, bases its component designs on the MVC framework.Model, View, and Controller are the three components that make up a GUI component according to the MVC paradigm, and each is important in establishing how the component behaves. The MVC design pattern encourages a clear separation of concerns and aids modularity, maintainability, and reusability in the design and development of the system by breaking the programme component into these discrete sections.

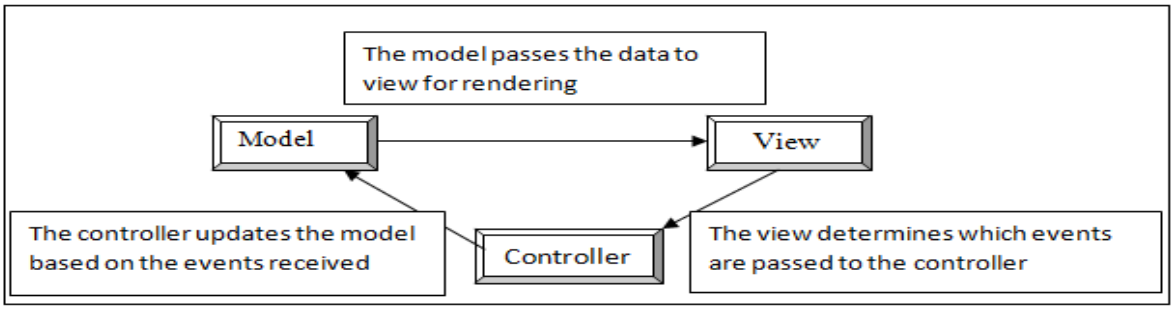


Fig 2-Communication through the MVC architecture

IX. PROPOSED ALGORITHM

a. KNN Algorithm:

K nearest neighbors is one of the simplest machine learning algorithms based on supervised learning techniques.The K-NN algorithm accepts the similarity between the new case/new data and the available cases and places the new case in the category most similar to the available categories.

- **Step-1:** Select the number K of the neighbors

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 4, April 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1104048 |

- **Step-2:** Calculate the Euclidean distance of K number of neighbors
- **Step-3:** Take the K nearest neighbors as per the calculated Euclidean distance.
- **Step-4:** Among these k neighbors, count the number of the data points in each category.
- **Step-5:** Assign the new data points to that category for which the number of the neighbor is maximum.
- **Step-6:** Our model is ready.

b. Random Forest:

An incredibly common supervised machine learning technique used for Classification and Regression issues in machine learning is called the Random Forest Algorithm. A forest is made up of many different types of trees, and the more trees there are, the more robust the forest will be. Similar to this, the accuracy and problem-solving capacity of a Random Forest Algorithm increase with the number of trees in the algorithm. In order to increase the dataset's predictive accuracy, a classifier called Random Forest uses many decision trees on different subsets of the input data. It is based on the idea of ensemble learning, which is the practice of integrating various classifiers to solve a challenging problem and enhance the model's performance.

- **Step-1:** Randomly select "k" features from total "m" features, where k << m
- **Step-2:** Among the "k" features, calculate the node "d" using the best split point.
- **Step-3:** Split the node into daughter nodes using the best split (see entropy and information gain above).
- **Step-4:** Repeat 1 to 3 steps until "l" number of nodes has been reached.
- **Step-5:** Build forest by repeating steps 1 to 4 for "n" number times to create "n" number of trees.

c. Logistic Regression Algorithm

Logistic regression is a statistical method for analyzing data that involves predicting the outcome of a categorical dependent variable based on one or more independent variables. It is commonly used in machine learning for binary classification tasks.

- **Step-1:** Draw the scatterplot. Look for 1) linear or non-linear pattern of the data and 2) deviations from the pattern (outliers). If the pattern is non-linear, consider a transformation. If there are outliers, you may consider removing them only IF there is a non-statistical reason to do so. (Are those individuals "different" than the rest of the sampled individuals?)
- **Step-2:** Fit the least-squares regression line to the data and check the assumptions of the model by looking at the Residual Plot (for constant standard deviation assumption) and normal probability plot (for normality assumption). If the assumptions of the model appear not to be met, a transformation may be necessary.
- **Step-3:** If necessary, transform the data and re-fit the least-squares regression line using the transformed data.
- **Step-4:** If a transformation was done, go back to step 1.
- **Step-5:** Once a "good-fitting" model is determined, write the equation of the least-squares regression line. Include the standard errors of the estimates, the estimate of σ, and R-squared.
- **Step-6:** Determine if the explanatory variable is a significant predictor of the response variable by performing a t-test or F-test. Include a confidence interval for the estimate of the regression coefficient (slope).

X. RESULT

Four alternative classifiers—Logistic Regression, Random Forest Classifier, Linear SVC, and KNeighbors Classifier—as well as two feature extraction techniques—count vectorization (CV) and term frequency-inverse document frequency (TF-IDF)—are compared in this study. To ascertain the effectiveness of various methods and algorithms within the framework of the study, their performance will be assessed and compared.This analytical study is illustrated in thefigurebelow.

By this Prediction we can say that the Random Forest will give more accuracy of 95%compare to all other algorithms

Algorithm	Expected Output	Actual Output
Random Forest	100%	95%
Linear SVC	100%	94%
Logistic Regression	100%	93%
KNeighbors	100%	85%

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 4, April 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1104048 |

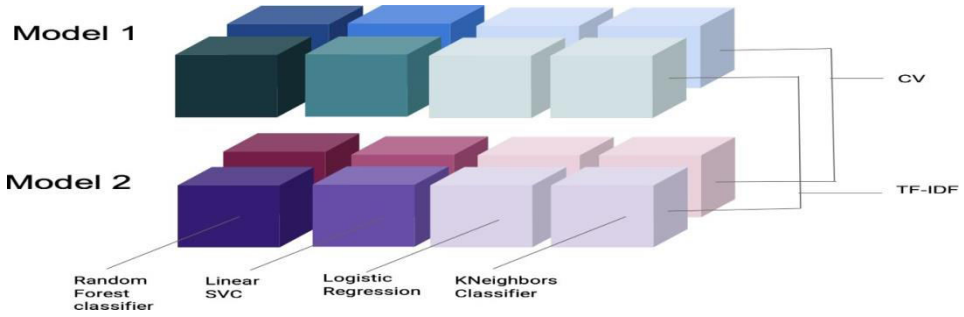


Fig 3: Comparative Analysis between two feature Selections methods

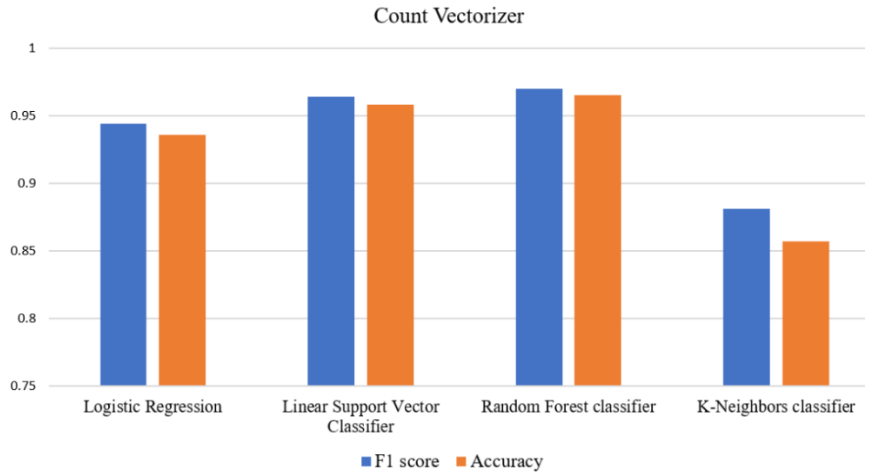


Fig 4: Comparison of Algorithms with count vectorizer

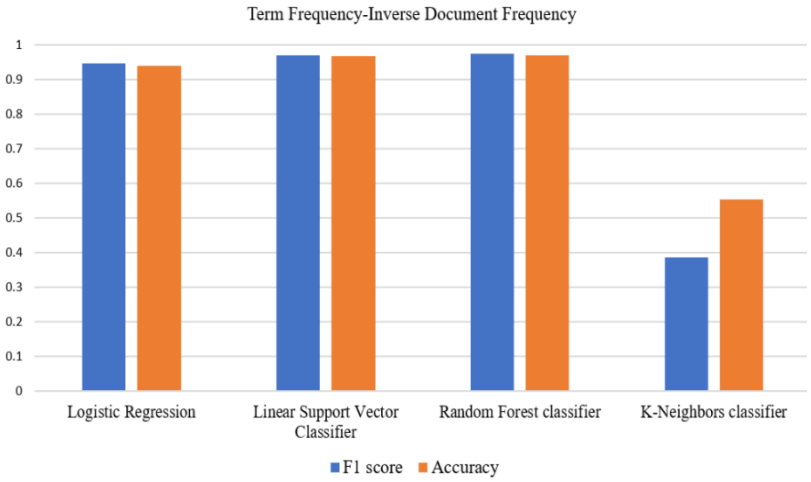


Fig 5: Comparison of Algorithms with Term Frequency Inverse Document frequency

XI. CONCLUSION AND FUTURE DISCUSSION

Our ground-breaking approach uses supervised machine learning and natural language processing to efficiently identify cyberbullying as well as related themes/categories like racism, sexism, physical abuse, vulgarity, and more. Our

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 4, April 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1104048 |

method includes a flexible hypothesis that may adjust to shifting linguistic trends over time to ensure reliable detection. Importantly, our method has produced results with great accuracy, with precision, recall, and F1 Score all going beyond 90%. We are continually attempting to increase accuracy levels as our study progresses. We also intend to investigate additional user and network factors that can affect cyberbullying detection in upcoming investigations.

REFERENCES

Here is a list of six research articles focused on detecting Android malware:

1. D.Yin,Z.Xue,andL.Hong,"DetectionofHarassmentonWeb2.0,"p.8,2019.
2. S. O. Sood, E. F. Churchill, and J. Antin, "Automatic identification of personal insults on socialnewssites,"J.Am.Soc.Inf.Sci.Technol.,vol.63,no.2,pp.270–285,Feb.2012,doi:10.1002/asi.21690.
3. A.Squicciarini,S.Rajtmajer,Y.Liu,andC.Griffin,"Identificationandcharacterizationofcyberbullyingdynamicsinanonlinesocialnetwork,"inProceedingsofthe2015IEEE/ACMInternational Conference on Advances in SocialNetworks Analysis andMining 2015 - ASONAM'15,Paris, France, 2015,pp. 280–285, doi: 10.1145/2808797.2809398.
4. V. S. Chavan and Shylaja S S, "Machine learning approach for detection of cyber-aggressivecomments by peers on social media network," in 2015 International Conference on Advances inComputing,Communicationsand Informatics(ICACCI), Kochi,India,Aug. 2015, pp. 2354–2358,doi:10.1109/ICACCI.2015.7275970.
5. "InternetArchiveSearch:collection:twitterstream."
6. https://archive.org/search.php?query=collection%3Atwitterstream&sort=publicdate&page=2(accessedJul. 23, 2020).

INNO SPACE
SJIF Scientific Journal Impact Factor

Impact Factor: 8.165

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details