www.ijbar.org
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86
https://doi.org/10.5281/zenodo.14066347

PREDICTIVE MODELS FOR CYBERBULLYING DETECTION ON TWITTER USING MACHINE LEARNING

[1]M.Mounika, [2]Vemundla Shirisha
[1]Assistant Professor,[2]MCA Student
Depatment Of MCA
Sree Chaitanya College of Engineering, Karimnagar

ABSTRACT

In order to automatically classify tweets into predetermined classes or categories, this paper presents a machine learning-based method for creating a twitter classifier. The suggested tweet classifier uses machine learning algorithms to evaluate tweet content and classify it into appropriate groups according to its semantic meaning and context, taking use of the enormous quantity of textual data accessible on social media sites like Twitter. To extract pertinent characteristics for categorisation, the development process preprocesses the twitter data using techniques like tokenisation, stemming, and stop word removal. To discover patterns and connections between tweet content and categories, a variety of machine learning models are trained on labelled twitter datasets, including Naive Bayes, Support Vector Machines (SVM), and Convolutional Neural Networks (CNN). Using parameters like accuracy, precision, recall, and F1-score on a test dataset, the tweet classifier's efficacy is assessed. Results from experiments show how reliable and effective the machine learning-based twitter classifier is at correctly classifying tweets, which offers important insights into subject recognition, sentiment analysis, and social media trends. All things considered, the suggested method provides a strong instrument for automating tweet categorisation tasks and gleaning valuable information from massive amounts of social media data.

I. INTRODUCTION

Social media networks such as Face book, Twitter, Flickr, and Instagram have become the preferred online platforms for interaction and socialization among people of all ages. While these platforms enable people to communicate and interact in previously unthinkable ways, they have also led to malevolent activities such as cyber-bullying. Cyber bullying is a type of psychological abuse with a significant impact on society. Cyber-bullying events have been increasing mostly among young people spending most of their time navigating between different social media platforms. Particularly, social media networks such as Twitter and Face book are prone to CB because of their popularity and the anonymity that the Internet provides to abusers. In India, for example, 14 percent of all harassment occurs on Face book and Twitter, with 37 percent of these incidents involving youngsters [1]. Moreover, cyber bullying might lead to serious mental issues and adverse mental health effects. Most

www.ijbar.org
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86
https://doi.org/10.5281/zenodo.14066347

suicides are due to the anxiety, depression, stress, and social and emotional difficulties from cyber-bullying events [2]_[4]. This motivates the need for an approach to identify cyber bullying in social media messages (e.g., posts, tweets, and comments).

In this article, we mainly focus on the problem of cyber bullying detection on the Twitter platform. As cyber bullying is becoming a prevalent problem in Twitter, the detection of cyber bullying events from tweets and provisioning preventive measures are the primary tasks in battling cyber bullying threats [5]. Therefore, there is a greater need to increase the research on social networks-based CB in order to get greater insights and aid in the development of effective tools and approaches to effectively combat cyber bullying problem [6]. Manually monitoring and controlling cyber bullying on Twitter platform is virtually impossible [7]. Furthermore, mining social media messages for cyber bullying detection is quite difficult. For example, Twitter messages are often brief, full of slang, and may include emojis, and gifs, which makes it impossible to deduce individuals' intentions and meanings purely from social media messages. Moreover, bullying can be difficult to detect if the bully uses strategies like sarcasm or passive-aggressiveness to conceal it. Despite the challenges that social media messages bring, cyber bullying detection on social media is an open and active research topic. Cyber bullying detection within the Twitter platform has largely been pursued through tweet classification and to a certain extent with topic modeling approaches. Text classification based on supervised machine learning (ML) models are commonly used for classifying tweets into bullying and non-bullying tweets [8]_[17]. Deep learning (DL) based classifiers have also been used for classifying tweets into bullying and non-bullying tweets [7], [18]_[22]. Supervised classifiers have low performance in case the class labels are unchangeable and are not relevant to the new events [23]. Also, it may be suitable only for a pre-determined collection of events, but it cannot successfully handle tweets that change on the _y. Topic modeling approaches have long been utilized as the medium to extract the vital topics from a set of data to form the patterns or classes in the complete dataset. Although the concept is similar, the general unsupervised topic models cannot be efficient for short texts, and hence specialized unsupervised short text topic models were employed [24]. These models effectively identify the trending topics from tweets and extract them for further processing. These models help in leveraging the bidirectional processing to extract meaningful topics. However, these unsupervised models require extensive training to obtain sufficient prior knowledge, which is not adequate in all cases [25]. Considering these limitations, an efficient tweet classification approach must be developed to bridge the gap between the classifier and the topic model so that the adaptability is significantly proficient.

www.ijbar.org
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86
https://doi.org/10.5281/zenodo.14066347

In this article, we propose a hybrid deep learning-based approach, called DEA-RNN, which automatically detects bullying from tweets. The DEA-RNN approach combines Elman type Recurrent Neural Networks (RNN) with an improved Dolphin Echolocation Algorithm (DEA) for _netuning the Elman RNN's parameters. DEA-RNN can handle the dynamic nature of short texts and can cope with the topic models for the effective extraction of trending topics. DEA-RNN outperformed the considered existing approaches in detecting cyber bullying on the Twitter platform in all scenarios and with various evaluation metrics.

The contributions of this article can be summarized as the following

_ Develop an improved optimization model of DEA for use to automatically tune the RNN parameters to enhance the performance;

_ Propose DEA-RNN by combining the Elman type RNN and the improved DEA for optimal classification of tweets;

_ A new Twitter dataset is collected based on cyber bullying keywords for evaluating the performance of DEA-RNN and the existing methods; and

_ The efficiency of DEA-RNN in recognizing and classifying cyber bullying tweets is assessed using Twitter datasets. The thorough experimental results reveal that DEA-RNN outperforms other competing models in terms of recall, precision, accuracy, F1 score, and specificity.

II. LITERATURE SURVEY

Mishna, M. Khoury-Kassabri, T. Gadalla, and J. Daciuk, ``Risk factors for involvement in cyber bullying: Victims, bullies and bully_victims,"

The purpose of the current study was to examine the frequency of cyber bullying among youth by distinguishing among the three categories of involvement in cyber bullying: victims, bullies, and bully–victims, to compare these to a fourth category of students who are not involved in the three categories of cyber bullying and to explore the factors that contribute to involvement in cyber bullying.

K. Miller, ``Cyberbullying and its consequences: How cyberbullying is contorting the minds of victims and bullies alike, and the law's limited available redress.

The use of digital and social media is growing every day as technology advances. People in the twenty-first century are growing up in a social media and internet-enabled society. Digital media offers a lot of opportunities, but people frequently tend to misuse them. On social networking sites, people spread anger toward a person. People are affected by cyberbullying in various ways. It has an impact on more than just health; numerous other factors put life in danger. Cyberbullying is a widespread modern phenomenon that people cannot completely avoid but can prevent. The author proposes a system for automatic cyberbullying detection and prevention using supervised machine learning. The system considers key characteristics of cyberbullying, such as the intention to harm,

www.ijbar.org
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86
https://doi.org/10.5281/zenodo.14066347

repeated behavior, and the use of abusive language. Support vector machines and logistic regression are employed to identify cyberbullying and related themes/categories such as race, physical, sexuality, and politics.

This proposed method offers a novel theory for the detection of cyberbullying: texting has evolved over time due to changes in context usage, and language. In the dataset that includes tweets, Support Vector Machine (SVM), Naïve Bayes, and Logistic Regression (LR) models were tested along with different Natural Language Processing methods. The accuracy of the system is improved by sentiment analysis, N-gram analysis, and other non-traditional feature extraction methods like Term Frequency-Inverse Document Frequency (TF-IDF) and profanity detection.

M. Vivolo-Kantor, B. N. Martell, K. M. Holland, and R. Westby, ``A systematic review and content analysis of bullying and cyber-bullying measurement strategies

Bullying has emerged as a behavior with deleterious effects on youth; however, prevalence estimates vary based on measurement strategies employed. We conducted a systematic review and content analysis of bullying measurement strategies to gain a better understanding of each strategy including behavioral content. Multiple online databases (i.e., PsychInfo, MedLine, ERIC) were searched to identify measurement strategies published between 1985 and 2012. Included measurement strategies assessed bullying behaviors, were administered to respondents with ages of 12 to 20, were administered in English, and included psychometric data. Each publication was coded independently by two study team members with a pre-set data extraction form, who subsequently met to discuss discrepancies. Forty-one measures were included in the review. A majority used differing terminology; student self-report as primary reporting method; and included verbal forms of bullying in item content. Eleven measures included a definition of bullying, and 13 used the term "bullying" in the measure. Very few definitions or measures captured components of bullying such as repetition, power imbalance, aggression, and intent to harm. Findings demonstrate general inconsistency in measurement strategies on a range of issues, thus, making comparing prevalence rates between measures difficult.

H. Sampasa-Kanyinga, P. Roumeliotis, and H. Xu, ``Associations between cyberbullying and school bullying victimization and suicidal ideation, plans and attempts among Canadian schoolchildren

The negative effects of peer aggression on mental health are key issues for public health. The purpose of this study was to examine the associations between cyberbullying and school bullying victimization with suicidal ideation, plans and attempts among middle and high school students, and to test whether these relationships were mediated by reports of depression. Methods: Data for this study are from the 2011 Eastern Ontario Youth Risk

www.ijbar.org
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86
https://doi.org/10.5281/zenodo.14066347

Behaviour Survey, which is a cross-sectional regional school-based survey that was conducted among students in selected Grade 7 to 12 classes (1658 girls, 1341 boys; mean6SD age: 14.361.8 years). Results: Victims of cyberbullying and school bullying incurred a significantly higher risk of suicidal ideation (cyberbullying: crude odds ratio, 95% confidence interval = 3.31, 2.16–5.07; school bullying: 3.48, 2.48–4.89), plans (cyberbullying: 2.79, 1.63–4.77; school bullying: 2.76, 2.20–3.45) and attempts (cyberbullying: 1.73, 1.26–2.38; school bullying: 1.64, 1.18–2.27) compared to those who had not encountered such threats. Results were similar when adjusting for sociodemographic characteristics, substance use, and sedentary activities. Mediation analyses indicated that depression fully mediated the relationship between cyberbullying victimization and each of the outcomes of suicidal ideation, plans and attempts. Depression also fully mediated the relationship between school bullying victimization and suicide attempts, but partially mediated the relationship between school bullying victimization and both suicidal ideation and plans. Conclusion: These findings support an association between both cyberbullying and school bullying victimization and risk of suicidal ideation, plans and attempts. The mediating role of depression on these links justifies the need for addressing depression among victims of both forms of bullying to prevent the risk of subsequent suicidal behaviours.

III. SYSTEM ANALYSIS AND SYSTEM DESIGN

Existing System

The existing system for building a tweet classifier using machine learning techniques involves the utilization of various algorithms and methodologies to analyze and classify tweets based on their content. Typically, this process begins with the collection of a labeled dataset containing tweets categorized into different classes or topics, trained on this dataset to learn patterns and relationships between features extracted from the text of tweets and their corresponding classes. Additionally, techniques such as feature engineering, dimensionality reduction, and sentiment analysis may be applied to enhance the accuracy and effectiveness of the classifier. While existing tweet classifiers provide valuable insights into public opinion, sentiment analysis, and topic modeling, they may encounter challenges with noisy data, domain-specific language, and evolving trends in social media content. Moreover, the performance of existing classifiers may vary depending on factors such as the size and quality of the training dataset, the choice of features, and the complexity of the classification task. Despite these challenges, tweet classifiers serve as valuable tools for analyzing social media data and extracting meaningful insights for various applications such as marketing, public opinion analysis, and trend detection.

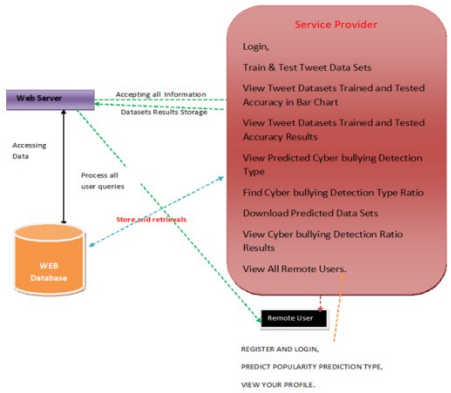
Proposed System

The proposed system for building a tweet classifier using machine learning aims to revolutionize the process of analyzing and categorizing tweets based on their content

www.ijbar.org
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86
https://doi.org/10.5281/zenodo.14066347

and sentiment. Leveraging state-of-the-art machine learning algorithms such as Naive Bayes, Support Vector Machines (SVM), and Neural Networks are then trained on this dataset to learn patterns and relationships between features extracted from the text of tweets and their corresponding classes and natural language processing techniques, the system will be capable of automatically classifying tweets into predefined categories such as positive, negative, or neutral sentiments, or into custom categories based on specific criteria defined by the user. By training the classifier on a diverse dataset of labeled tweets, the system will learn to recognize patterns and relationships in the text data, enabling accurate and efficient classification of new tweets. Additionally, the proposed system will offer features such as real-time data ingestion, preprocessing, and model training, allowing for continuous learning and adaptation to evolving trends and topics on social media platforms. Through a user-friendly interface, users will be able to interact with the system, customize classification criteria, and visualize classification results through intuitive dashboards and reports. Overall, the proposed machine learning-based tweet classifier promises to provide valuable insights into social media conversations, facilitate sentiment analysis, and support decision-making processes for businesses, marketers, and researchers.

SYSTEM ARCHITECTURE



IV. SYSTEM IMPLEMENTATION

Modules Description

Service Provider

In this module, the Service Provider has to login by using valid user name and password. After login successful he can do some operations such as

Login, Train & Test Tweet Data Sets, View Tweet Datasets Trained and Tested Accuracy in Bar Chart, View Tweet Datasets Trained and Tested Accuracy Results, View Predicted Cyber bullying Detection Type, Find Cyber bullying Detection Type Ratio, Download Predicted Data Sets, View Cyber bullying Detection Ratio Results, View All Remote Users.

View and Authorize Users

In this module, the admin can view the list of users who all registered. In this, the admin can view the user's details such as, user name, email, address and admin authorizes the users.

Remote User

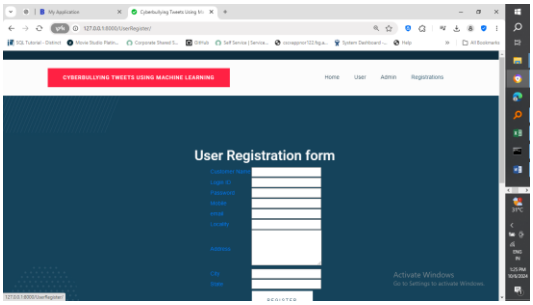
In this module, there are n numbers of users are present. User should register before doing any operations. Once user registers, their details will be stored to the database.

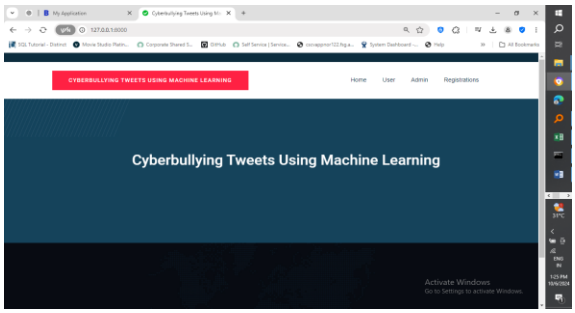
www.ijbar.org
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86
https://doi.org/10.5281/zenodo.14066347

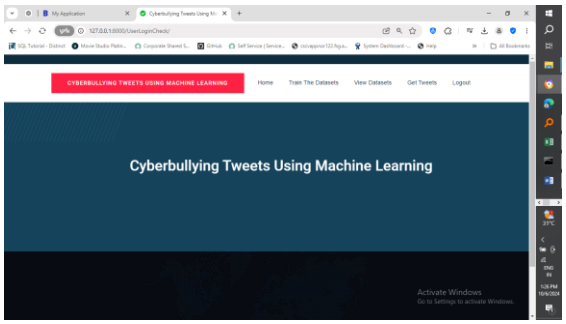
After registration successful, he has to login by using authorized user name and password. Once Login is successful user will do some operations like REGISTER AND LOGIN, PREDICT CYBERBULLYING TYPE, VIEW YOUR PROFILE.

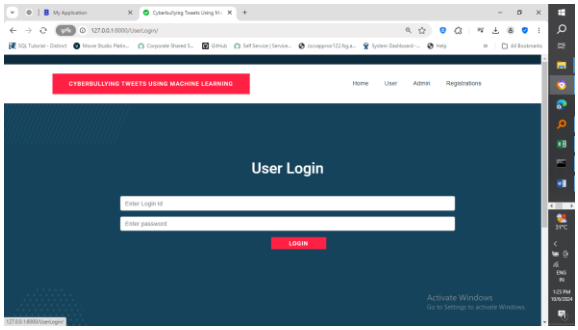
V. SCREEN SHOTS

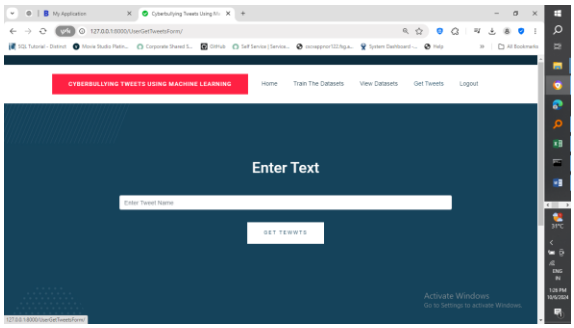
http://127.0.0.1:8000/

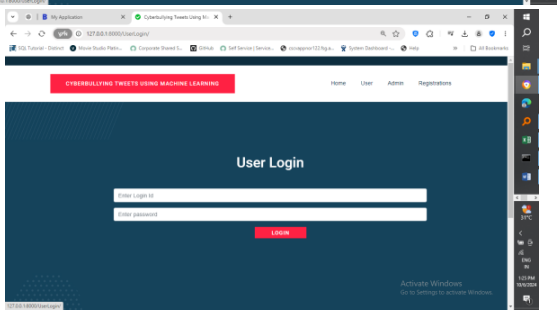


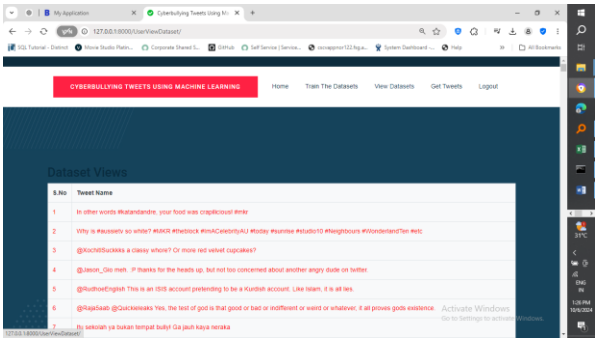




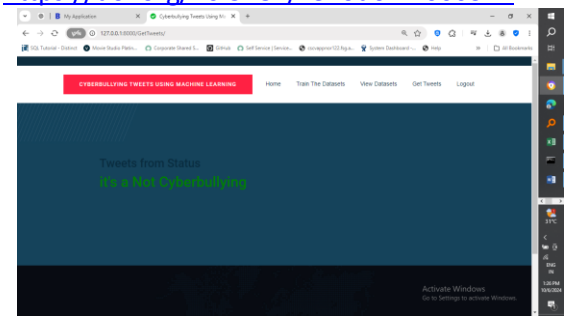








www.ijbar.org
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86
https://doi.org/10.5281/zenodo.14066347



VI. CONCLUSION

Our survey's results indicate that traditional machine learning algorithms are unable to handle the massive volumes of data produced by Web 4.0 and that cyberbullying content cannot be efficiently identified. Deep learning techniques like stacking auto-encoders, convolutional neural networks, and deep recurrent neural networks have recently attracted the attention of several academics. Future study may focus on using these deep learning algorithms to accurately identify cyberbullying on social media. Additionally, our cyberbullying detection technology is based on binary classification (bullying or non-bullying), therefore our future study may also focus on multi-class categorisation.

REFERENCES

[1] F. Mishna, M. Khoury-Kassabri, T. Gadalla, and J. Daciuk, ``Risk factors for involvement in cyber bullying: Victims, bullies and bully_victims," *Children Youth Services Rev.*, vol. 34, no. 1, pp. 63_70, Jan. 2012, doi:10.1016/j.childyouth.2011.08.032.

[2] K. Miller, ``Cyberbullying and its consequences: How cyberbullying is contorting the minds of victims and bullies alike, and the law's limited available redress," *Southern California Interdiscipl. Law J.*, vol. 26, no. 2, p. 379, 2016.

[3] A. M. Vivolo-Kantor, B. N. Martell, K. M. Holland, and R. Westby, ``A systematic review and content analysis of bullying and cyber-bullying measurement strategies," *Aggression Violent Behav.*, vol. 19, no. 4, pp. 423_434, Jul. 2014, doi: 10.1016/j.avb.2014.06.008.

[4] H. Sampasa-Kanyinga, P. Roumeliotis, and H. Xu, ``Associations between cyberbullying and school bullying victimization and suicidal ideation, plans and attempts among Canadian schoolchildren," *PLoS ONE*, vol. 9, no. 7, Jul. 2014, Art. no. e102145, doi: 10.1371/journal.pone.0102145.

[5] M. Dadvar, D. Trieschnigg, R. Ordelman, and F. de Jong, ``Improving cyberbullying detection with user context," in *Proc. Eur. Conf. Inf. Retr.*,in Lecture Notes in Computer Science: Including Subseries Lecture Notes in Arti_cial Intelligence and Lecture Notes in Bioinformatics, vol. 7814,
2013, pp. 693_696.

[6] A. S. Srinath, H. Johnson, G. G. Dagher, and M. Long, ``BullyNet: Unmasking cyberbullies on social networks," *IEEE Trans. Computat. Social Syst.*, vol. 8, no. 2, pp. 332_344, Apr. 2021, doi: 10.1109/TCSS.2021.3049232.

[7] A. Agarwal, A. S. Chivukula, M. H. Bhuyan, T. Jan, B. Narayan, and M. Prasad, ``Identi_cation and classi_cation of cyberbullying posts: A recurrent neural network approach using under-sampling and class weighting," in *Neural Information*

www.ijbar.org
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86
https://doi.org/10.5281/zenodo.14066347

Processing (Communications in Computer and Information Science), vol. 1333, H. Yang, K. Pasupa, A. C.-S. Leung, J. T. Kwok, J. H. Chan, and I. King, Eds. Cham, Switzerland: Springer, 2020, pp. 113_120.

[8] Z. L. Chia, M. Ptaszynski, F. Masui, G. Leliwa, and M. Wroczynski,``Machine learning and feature engineering-based study into sarcasm and irony classi_cation with application to cyberbullying detection,"*Inf. Process. Manage.*, vol. 58, no. 4, Jul. 2021, Art. no. 102600, doi: 10.1016/j.ipm.2021.102600.

[9] N. Yuvaraj, K. Srihari, G. Dhiman, K. Somasundaram, A. Sharma,S. Rajeskannan, M. Soni, G. S. Gaba, M. A. AlZain, and M. Masud,``Nature-inspired-based approach for automated cyberbullying classi_cation on multimedia social networking," *Math. Problems Eng.*, vol. 2021, pp. 1_12, Feb. 2021, doi: 10.1155/2021/6644652.

[10] B. A. Talpur and D. O'Sullivan, ``Multi-class imbalance in text classification: A feature engineering approach to detect cyberbullying in Twitter," *Informatics*, vol. 7, no. 4, p. 52, Nov. 2020, doi: 10.3390/informatics7040052.

[11] A. Muneer and S. M. Fati, ``A comparative analysis of machine learning techniques for cyberbullying detection on Twitter," *Futur. Internet*, vol. 12, no. 11, pp. 1_21, 2020, doi: 10.3390/_12110187.

[12] R. R. Dalvi, S. B. Chavan, and A. Halbe, ``Detecting a Twitter cyberbullying using machine learning," *Ann. Romanian Soc. Cell Biol.*, vol. 25, no. 4, pp. 16307_16315, 2021.

[13] R. Zhao, A. Zhou, and K. Mao, ``Automatic detection of cyberbullying on social networks based on bullying features," in *Proc. 17th Int. Conf. Distrib. Comput. Netw.*, Jan. 2016, pp. 1_6, doi: 10.1145/2833312.2849567.

[14] L. Cheng, J. Li, Y. N. Silva, D. L. Hall, and H. Liu, ``XBully: Cyberbullying detection within a multi-modal context," in *Proc. 12th ACM Int. Conf. Web Search Data Mining*, Jan. 2019, pp. 339_347, doi: 10.1145/3289600.3291037.

[15] K. Reynolds, A. Kontostathis, and L. Edwards, ``Using machine learning to detect cyberbullying," in *Proc. 10th Int. Conf. Mach. Learn. Appl. Workshops (ICMLA)*, vol. 2, Dec. 2011, pp. 241_244, doi:10.1109/ICMLA.2011.152.

[16] S. Agrawal and A. Awekar, ``Deep learning for detecting cyberbullying across multiple social media platforms," in *Advances in Information Retrieval* (Lecture Notes in Computer Science), vol. 10772, G. Pasi, B. Piwowarski, L. Azzopardi, and A. Hanbury, Eds. Cham, Switzerland: Springer, 2018, pp. 141_153. [17] R. I. Ra_q, H. Hosseinmardi, R. Han, Q. Lv, S. Mishra, and S. A. Mattson, ``Careful what you share in six seconds: Detecting cyberbullying instances in vine," in *Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining (ASONAM)*, Aug. 2015, pp. 617_622, doi: 10.1145/2808797.2809381.

[18] N. Yuvaraj, V. Chang, B. Gobinathan, A. Pinagapani, S. Kannan, G. Dhiman, and A. R. Rajan, ``Automatic detection of cyberbullying using multi-feature based arti_cial intelligence with deep decision tree

ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86
https://doi.org/10.5281/zenodo.14066347
classification," *Comput. Electr. Eng.*, vol. 92, Jun. 2021, Art. no. 107186, doi:10.1016/j.compeleceng.2021.107186.

[19] A. Al-Hassan and H. Al-Dossari, ``Detection of hate speech in Arabic tweets using deep learning," *Multimedia Syst.*, Jan. 2021, doi:10.1007/s00530-020-00742-w.

[20] Y. Fang, S. Yang, B. Zhao, and C. Huang, ``Cyberbullying detection in social networks using bi-GRU with self-attention mechanism," *Information*, vol. 12, no. 4, p. 171, Apr. 2021, doi: 10.3390/info12040171.