



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

IJIRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 12, Issue 9, September 2023

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.423

9940 572 462

6381 907 438

ijirset@gmail.com

www.ijirset.com

Cyber Aggression Detection on COVID-19 Data

Soorya C G, Dr S Jothi

P.G. Student, Department of Computer Engineering, KCG College of Technology, Chennai, Tamilnadu, India

Professor, Department of Computer Engineering, KCG College of Technology, Chennai, Tamilnadu, India

ABSTRACT: A subtype of cyber-aggression that has become a major issue in today's information-driven culture is cyber bullying. The internet's accessibility has made it easier for ideas and knowledge to spread in the digital age, thus heightening societal awareness. Nevertheless, this advancement brings its own set of challenges. The surge in online platforms has opened avenues for propagators of misinformation to entice individuals with their misleading viewpoints and fabricated news. The urgency to counteract such misinformation and hate speech has escalated during the COVID-19 pandemic. This study undertakes a comprehensive examination of a spectrum of scholarly articles that address analogous problem statements, including topics like topic extraction, sentiment analysis, and fake news. The primary objective is to harness the insights from established research to bolster the ongoing exploration of COVID-19-related datasets. Employing Natural Language Processing (NLP) techniques for data preprocessing, this work leverages a Convolutional Neural Network (CNN) as well as Long Short-Term Memory (LSTM) component-based Hybrid Deep Learning model, to discern and classify instances of cyber aggression in textual data. Remarkably, the Hybrid Deep Learning model, specifically the CLSTM model, achieves an impressive accuracy of 99%, whereas the Random Forest model exhibits a 70% accuracy rate.

KEYWORDS: *Cyber bullying, COVID-19 pandemic, CLSTM model, Random Forest model.*

I. INTRODUCTION

The use of various Natural Language Processing (NLP) technologies to combat the rise of rumors, fake news, and even harmful and nasty comments has been the subject of numerous papers and technological developments. Globally, the ongoing COVID-19 epidemic has had an impact on people's life. The first COVID-19 cases, also known as SARS-CoV-2, were found in Wuhan, China, in December 2019. Initially confined to China or those recently visiting the country, the disease quickly spread. The COVID-19 pandemic was formally proclaimed by the WHO on March 11 signifying its broad societal impact beyond just a health crisis. The pandemic necessitated significant changes in how governments, healthcare professionals, and the general population approach daily activities while grappling with the virus. During such challenging times, the ability to authenticate the accuracy of tweets or articles proves invaluable. Text recognition can be used by NLP to identify trends in bogus news allowing the identification of articles employing similar language. As COVID-19 spread globally, its effects became palpable. Lockdowns were implemented to curb contagion rates, driven by widespread awareness fueled by pervasive communication. However, this platform for free expression also ushered in a surge of misinformation and hate speech. In essence, the convergence of technological innovation and global health challenges underscores the importance of NLP in combating misinformation and hate speech during the COVID-19 pandemic.

II. RELATED WORK

In order to take action by filtering out unpleasant comments or giving counsel to people affected, Raisi [1] has presented a technique to identify offensive statements on social networks. They used abusive tweets and Ask.fm comments to train their algorithm. Reynolds [2] introduced a system for identifying cyberbullying on the Formspring social network by recognizing violence patterns in user posts, analyzing offensive language, and assigning a threat level ranking. This system used J48 decision trees to obtain an accuracy of 81.7%. Ptaszynski [3] proposed the creation of a web application in Japan for parents and school employees to identify inappropriate material on unofficial secondary sites in order to submit instances of cyberbullying to federal authorities. This work used SVMs and had a 79.9% accuracy rate. In another study [4], 3915 messages that were published on Formspring were analyzed for offensive language. The accuracy in this study was only 58.5%. This research resulted in the creation of a system for detecting cyberbullying in YouTube video comments, including identifying the message's topic (e.g., sexuality, race, intelligence), with an overall success rate of 66.7% using SVMs.

Similarly, Nandhini [5] suggested a method to identify cyberbullying activities & classify them using a fuzzy categorization rule into subcategories including flaming, harassment, racism, & terrorism. However, the results' accuracy was only about 40% accurate, despite the classifier's efficiency being increased to 90% by a number of rules. A violent, intentional act committed online towards a victim who is powerless to defend themselves is known as cyberbullying [6]. A key distinction from traditional bullying is the perpetrator's intent to hurt the victim's feelings [7]. This includes an imbalance of power and repetitive contact, which are unique characteristics of cyberbullying. Cyberbullying may entail complications like the repeating of aggression & power imbalances on an online context, even though some people think of it as an electronic version of face-to-face bullying [8].

The relationship between cyberbullying and anonymity in online social networks has been extensively explored by Hosseinmardi et al. [9]. Ask.fm is a semi-anonymous online social network, allowing users to hide their identity when posting questions or comments on profiles. Andriansyah et al. [10] used Support Vector Machine (SVM) research to categorize comments on Instagram that contained cyberbullying. They used comments from Indonesian celebrities' accounts, namely Karin Novilda and Samuel Alexandar, with 1053 comments for training and 34 for testing. They created a text term matrix with R language to develop an SVM model, achieving an accuracy of 79.41% for identifying cyberbullying comments.

Dataset details: With social media's widespread use across all age groups, cyberbullying, facilitated by online anonymity, poses a pervasive threat. UNICEF warned of increased risks during the COVID-19 pandemic, and this dataset comprises 47,000 tweets categorized by cyberbullying class (age, ethnicity, gender, religion, other, non-cyberbullying). The dataset is balanced with approximately 8,000 instances for each class.

Dataset link:

<https://www.kaggle.com/datasets/andrewmvd/cyberbullying-classification>

Deep Learning

A subset of deep learning is machine learning, especially within neural networks with three layers or more. These neural networks allow computers to "learn" from enormous datasets, despite their efforts to emulate human brain processes frequently falling short. The inclusion of additional hidden layers can aid in fine-tuning and improving accuracy, even when a single-layer neural network provides approximate predictions. Deep learning is essential for enabling a variety of artificially intelligent (AI) applications & services, boosting automation by carrying out both mental and physical tasks independently. Such products and services as virtual assistants, controlled by voice TV remote controls, as well as credit card fraud prevention are smoothly connected with this technology. It powers not only innovative technological advancements like self-driving cars but also groundbreaking developments in various fields.

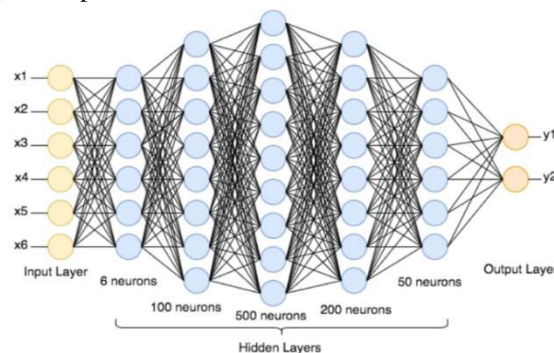


Fig 1: Architecture of Deep Learning

III. PROPOSED METHOD

System architecture planning identifies the WebApp's overall hypermedia structure. The goals set for a WebApp, the information to be displayed, the users whom will visit, & the specified navigation philosophy all have an impact on the architecture design.

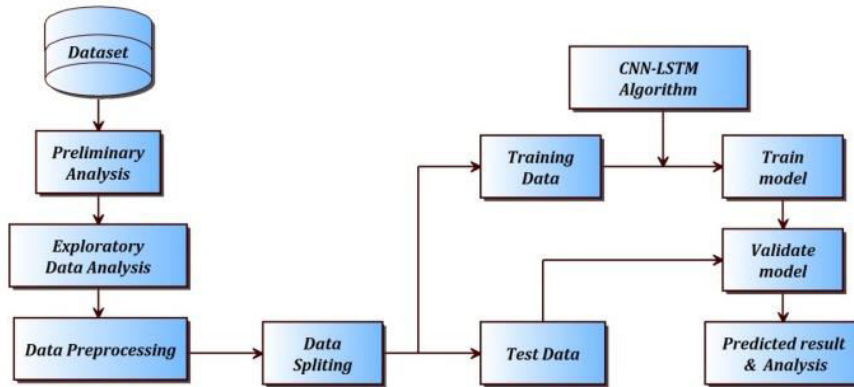


Fig2: Proposed System Architecture

The structuring of content objects for presentation & navigation is the subject of content architecture. The architecture of a web application describes how it is set up to handle user interaction, carry out internal processing activities, facilitate navigation, & present content. The definition of WebApp architecture takes into account the creation environment in which the application will be used.

1. One Dimensional Convolution Neural Network (CNN)

A One-Dimensional Convolutional Neural Network (1D CNN) is a specialized neural architecture tailored for processing one-dimensional data sequences like time series, audio, or text. Unlike traditional image-focused CNNs, 1D CNNs employ one-dimensional convolutional layers to detect patterns within sequential data. They utilize features like convolutional and pooling layers to automatically extract relevant information, making them ideal for tasks like text analysis, speech recognition, and financial forecasting. These networks offer translation invariance, recognizing patterns regardless of their position in the sequence. Widely applied in natural language processing, audio processing, and more, 1D CNNs excel in automated feature learning and pattern recognition for sequential data analysis.

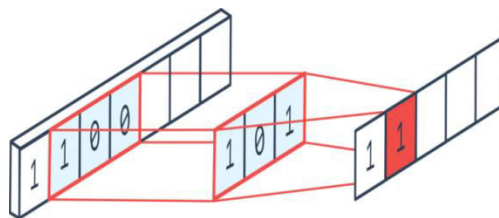


Fig 3: One Dimensional Convolution using a 3x1 stride and a 1x1 kernel.

2. LSTM Networks

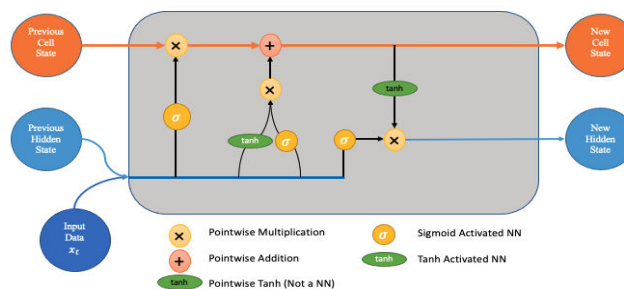


Fig 4: LSTM Model Diagram



This post delves into Long Short-Term Memory (LSTM) networks, which are crucial for tackling the challenge of long-term dependencies often faced in RNNs, or recurrent neural networks. Because LSTMs have feedback connections instead of just feed-forward connections, they can handle whole data sequences like time series, with a broader contextual understanding. They excel in handling time-varying data where each month's figures depend on the previous ones. LSTMs learn to capture patterns across extended time intervals, making them invaluable for tasks involving text, audio, and more. LSTMs operate through gates, including input and storage gates, managing the flow of information. Understanding these gates is fundamental to comprehending LSTM functionality and their role in handling long-range dependencies in sequential data.

3. Random Forest Classification

A potent machine learning method used for either classification & regression tasks are random forest classification. It works by building several decision trees during training, each tree offering a prediction. Then, it combines these predictions to produce a more accurate and robust result. The "random" aspect refers to the random selection of features and data points for each tree, reducing the risk of overfitting and enhancing generalization. Random Forests are highly versatile, perform well with large datasets, and are less prone to bias. They are widely employed in applications such as image classification, medical diagnosis, and fraud detection, offering a dependable approach to complex classification challenges.

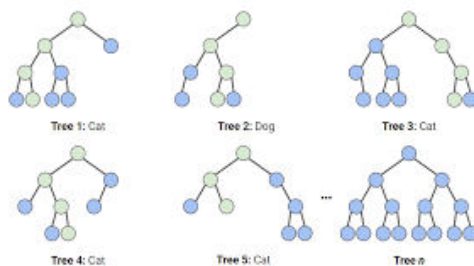


Fig 5: Random Forest Model Diagram

IV. RESULT ANALYSIS

The classification report offers an extensive evaluation of a classification model's performance, measured using analytical criteria for each target class, including recall, precision, F1-score, as well as contributions. These metrics collectively summarize the model's accuracy in correctly categorizing different groups. They hold significant value in assessing the model's effectiveness, precision, and its capability to accurately classify instances within numerous classes. The classification report analysis offers insightful information about the model's performance aiding in informed decisions about its suitability for a specific classification task.

Precision (P): Precision, often denoted as $TP / (TP + FP)$, assesses the model's accuracy in predicting positive outcomes. It gauges the model's ability to correctly identify relevant instances.

$$Precision = TP / (TP + FP)$$

When compared to all genuine positives & false negatives, the recall rate, also known as the genuine positive rate, calculates the percentage of actual positives that were correctly identified. This measure evaluates how well the system can spot promising advancements.

$$Recall = TP / (TP + FN)$$

F1-score: The F1 score represents a harmonious blend of precision and recall, striking a balance between these two metrics. It offers a way to find equilibrium between recall and precision in a performance evaluation.

$$F1\text{-score} = 2 * (Precision * Recall) / (Precision + Recall)$$

Support: Support denotes the count of samples within each target class.



CLSTM				
	Precision	recall	f1-score	support
other_cyberbullying	1	0.99	1	604
age	0.98	0.99	0.99	586
religion	0.99	0.98	0.99	588
gender	0.97	0.99	0.98	602
ethnicity	1	1	1	589
RANDOM FOREST CLASSIFICATION				
other_cyberbullying	0.88	0.96	0.92	603
age	0.68	0.63	0.66	578
religion	0.6	0.52	0.56	575
gender	0.63	0.7	0.66	567
ethnicity	0.66	0.65	0.65	646

Fig 6: Performance Metrics Comparison

In the table summarizing the performance metrics, the CLSTM model demonstrates strong classification results with high precision, recall, and F1-score across multiple target classes. It specifically achieves precision of 1.0, recall of 0.99, & an F1-score of 1.0 for the "other_cyberbullying" class, suggesting nearly flawless performance. Similarly, it exhibits excellent performance for the "age," "religion," "gender," and "ethnicity" classes, with F1-scores ranging from 0.98 to 1.0. In contrast, the Random Forest Classification model shows lower performance metrics across all target classes. For "other_cyberbullying," It has an F1-score of 0.92, a precision of 0.88, and a recall of 0.96. The F1-scores for the other classes range from 0.56 to 0.66, indicating lower overall accuracy compared to the CLSTM model. Notably, the CLSTM model outperforms the Random Forest Classification model significantly, with an accuracy of 99% compared to 70%, highlighting its superior classification capabilities.

Confusion matrix

A classification model's performance can be evaluated by comparing its forecasts with the dataset's actual values using the matrix of confusion, known as a square matrix. The confusion matrix's rows indicate the true or actual classes, while its columns display the anticipated classes.

A. Convolution Long Short Term Memory (CLSTM)

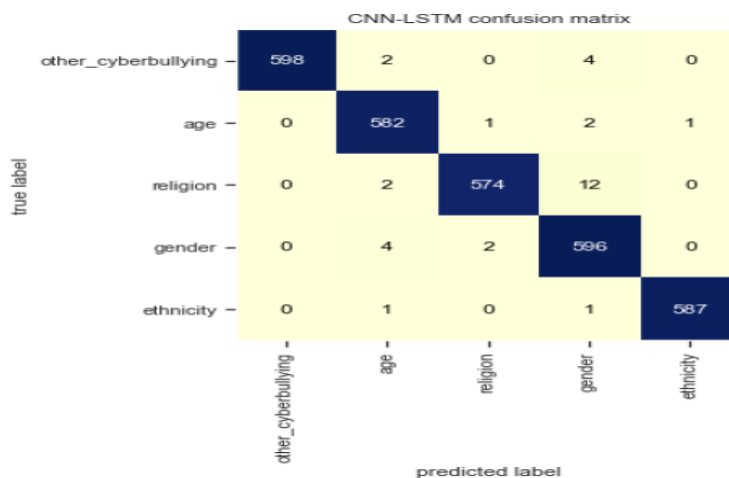


Fig 7: Confusion Matrix for CLSTM



Fig 6 displays the confusion matrix for the CLSTM model, illustrating its exceptional performance. This model attains an outstanding accuracy rate of 99.99%.

B. Random Forest Classification

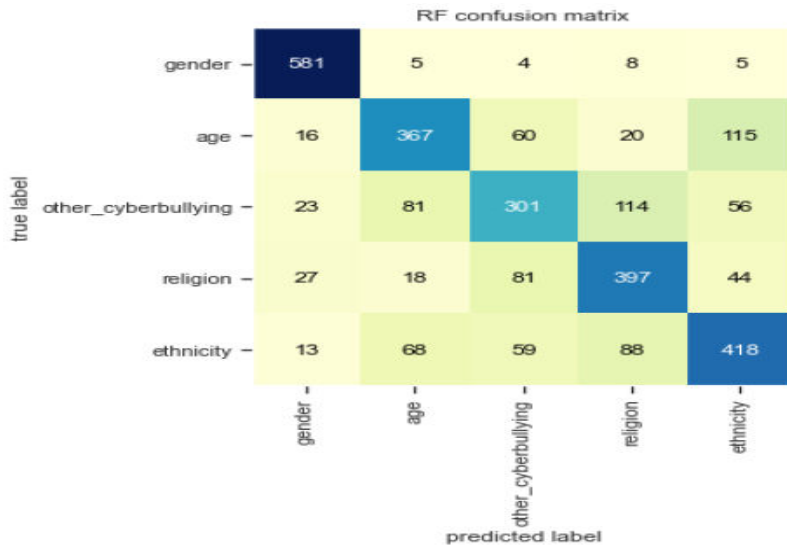


Fig 8: Confusion Matrix for Random Forest Classification

Fig. 7 illustrates the Random Forest Classification model's performance using a confusion chart, revealing an accuracy rate of 70%.

S.No	Model	Model Acc. %
1	CLSTM	0.99
2	Random forest	0.70

Table 1: Accuracy Comparison

It clearly shows CLSTM performs better than the Random forest model. The CLSTM model gives 99% accuracy whereas the Random forest model gives 70% accuracy. Random forest model gives 70%. CLSTM gives 99%. CLSTM overall accuracy is 99% and gives good precision for all classifications, so we can conclude CLSTM is performing better than Random forest.

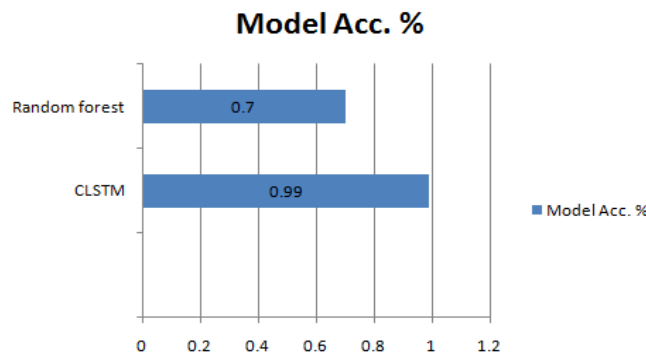
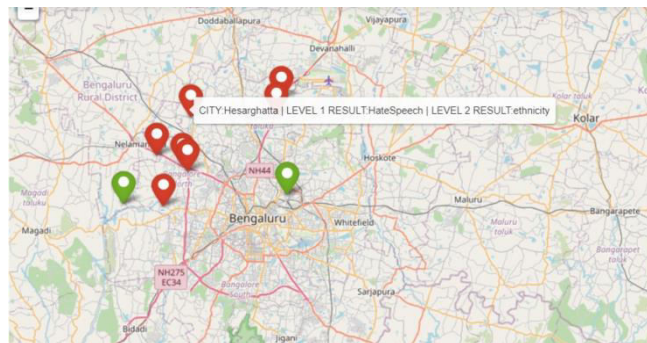


Fig9: Accuracy Comparison graph

V. MAP OUTPUT

	City	Lat	Lng	Text
0	Pakri	25.5876	85.1580	Been giving her the silent treatment, it was s...
1	Hunasamaranhalli	13.1435	77.6200	Dumb ass niggers with no heart ' fuck ya'll ...
2	Hesarghatta	13.1391	77.4783	Black is a color . African American is a Cultu...
3	Bommayapālaiyam	11.9922	79.8499	Noo fuck youu !!! & Stay Mad with your mad...
4	Gundūr	10.7339	78.7184	@KhaledHamaki He beheaded 600 Jewish prisoners...
5	Punādih	25.5484	85.2649	So attacking children because they are born wh...
6	Harilādih	23.6300	86.3500	@alaan_wtf ese tambien lo deberia de estar vie...
7	Alāwalpur	25.4958	85.2021	@TripleCeez004 woot
8	Mādnāikanhalli	13.0626	77.4642	the stupidest thing in the world is when a str...
9	Kādiganahalli	13.1687	77.6283	@CHEEEKABOO LOL I don't give a fuck you dumb b...

A dataframe is input to the system. After necessary natural language processing (NLP) processing is applied to the text column, the samples are sent to the model for prediction. The system can detect two levels of cyberbullying. The dataframe also includes additional information such as city, latitude, and longitude. This information is used to annotate a geomap. Green points on the map indicate normal targets, while red points indicate cyberbullying targets. Geo map output showed below



VI. CONCLUSION

In conclusion, this study delves into the pervasive issue of cyberbullying, a growing concern in our digitally connected society. With the internet's widespread accessibility, the dissemination of information has reached unprecedented levels, but it has also given rise to challenges like misinformation and hate speech. The COVID-19 pandemic has intensified the need to combat such problems, making this research timely and relevant. By analyzing a range of scholarly articles addressing related issues such as fake news and sentiment classification, this study seeks to harness insights from existing research to address COVID-19-related datasets effectively. Convolutional Neural Network (CNN) & Long Short-Term Memory (LSTM) components of a Hybrid Deep Learning model, along with Natural Language Processing (NLP) methods, this work successfully identifies and classifies instances of cyber aggression in textual data. Notably, the Hybrid Deep Learning model, particularly the CLSTM model, stands out with an impressive accuracy rate of 99%, showcasing its robustness in addressing cyberbullying. In contrast, the Random Forest model achieves a 70% accuracy rate. These findings underscore the potential of advanced NLP techniques and deep learning models, like CLSTM, in combating cyberbullying and related issues in the digital age.

Future work

Future research in this domain should encompass several key areas. Firstly, expanding the dataset to include a more extensive and diverse range of cyberbullying instances, particularly those related to emerging issues, can improve model generalization. Incorporating multimedia elements, like images and videos, alongside text analysis can provide a more holistic view of cyberbullying. Developing real-time monitoring systems for proactive intervention and enhancing the explainability of models are critical. Additionally, a user-centric approach focusing on victim support is essential. Adapting models to evolving online platforms, cross-cultural analysis, ethical considerations, collaboration with online platforms, and conducting



longitudinal studies are promising directions to ensure the effectiveness and responsible use of cyberbullying detection tools in the future.

REFERENCES

- [1] Raisi, E.; Huang, B. Cyberbullying identification using participant-vocabulary consistency. arXiv 2016, arXiv:1606.08084.
- [2] Reynolds, K.; Kontostathis, A.; Edwards, L. Using machine learning to detect cyberbullying. In Proceedings of the 2011 10th International Conference on Machine learning and applications and workshops (ICMLA), Honolulu, HI, USA, 18–21 December 2011; Volume 2, pp. 241–244.
- [3] Ptaszynski, M.; Dybala, P.; Matsuba, T.; Masui, F.; Rzepka, R.; Araki, K. Machine learning and affect analysis against cyber-bullying. In Proceedings of the 36th AISB, Leicester, UK, 29 March–1 April 2010; pp. 7–16.
- [4] Dinakar, K.; Reichart, R.; Lieberman, H. Modeling the detection of Textual Cyberbullying. Soc. Mob. Web 2011, 11, 11–17.
- [5] Nandhini, B.S.; Sheeba, J. Online social network bullying detection using intelligence techniques, Procedia Comput. Sci. 2015, 45, 485–492.
- [6] P. K. Smith, J. Mahdavi, M. Carvalho, S. Fisher, S. Russell, and N. Tippett, “Cyberbullying: its nature and impact in secondary school pupils,” J. Child Psychol. Psychiatry, vol. 49, no. 4, pp. 376–385, Apr.
- [7] H. Vandebosch and K. Van Cleemput, “Defining Cyberbullying: A Qualitative Research into the Perceptions of Youngsters,” CyberPsychologyBehav., vol. 11, no. 4, pp. 499–503, Aug.
- [8] R. M. Kowalski, S. Limber, and P. W. Agatston, Cyberbullying: bullying in the digital age. Wiley-Blackwell, 2012.
- [9] D. Maher, “Cyberbullying: an ethnographic case study of one Australian upper primary school class,” Youth Stud. Aust., vol. 27, no. 4, pp. 50–58.
- [10] H. Hosseinmardi, A. Ghasemianlangroodi, R. Han, Q. Lv, and S. Mishra, “Towards understanding cyberbullying behavior in a semianonymous social network,” in 2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2014), 2014, pp. 244–252.



INNO SPACE
SJIF Scientific Journal Impact Factor
Impact Factor: 8.423



ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com



www.ijirset.com

Scan to save the contact details