ISSN (Online): 2583-7052

International Journal of Advanced Trends in Engineering and Management (IJATEM)
pp. 132-143

Article Title: **Detection of Cyber bullying on Social Media using Deep Learning Algorithms**

Detection of Cyber bullying on Social Media using Deep Learning Algorithms

[1]A. Afroz Abbas, [2]B. Haritha, [3]K. Kiran, [4]D. Gowtham, [5]K. Hemanth,
[6]K. Chand Basha, [7]K. Purushotham

[1,2,3,4,5,6]UG Students, Department of Electronics and Communication Engineering, Sri Venkatesa Perumal College of Engineering and Technology, Puttur, AP, India.

[7]Assistant Professor, Department of Electronics and Communication Engineering, Sri Venkatesa Perumal College of Engineering and Technology, Puttur, AP, India.

ABSTRACT

With the proliferation of Internet usage in contemporary society, a vast amount of data has been generated. However, alongside its benefits, the cyber world presents its own array of challenges, among which cyberbullying stands out as a significant issue. Cyberbullying, constituting an online form of crime, encompasses various criminal activities facilitated through the Internet, computers, mobile phones, and other electronic devices. Previous research in detecting cyberbullying has encountered limitations, including data unavailability, hidden identities of perpetrators, and victims' privacy concerns. To address these constraints, this paper proposes an effective text mining approach utilizing machine learning algorithms to actively identify bullying texts. Unlike prior studies focusing solely on textual features, this study incorporates three distinct types of features: textual, behavioral, and demographic. Textual features encompass intimidating language indicative of potential cyberbullying outcomes. Meanwhile, behavioral features are derived from observed online actions, and demographic features include age, gender, and location information extracted from the dataset. By integrating these multifaceted features, the proposed method aims to enhance the accuracy and robustness of cyberbullying detection on social media platforms.

Keywords: Neural Network, Convolutional Neural Network, NLP, BERT

1. Introduction

Online apps for socializing are a place where we can express our thoughts and opinions and can even share our personal lives with another person [1]. We can access social media with the help of internet connection in our phones, laptops, PCs, tablets etc. The most well-known online media incorporates Facebook1, Twitter2, Instagram3, Tik-Tok4, etc. These days, web-based media is engaged with various areas like Coaching [2], Entrepreneurship [3], and furthermore for the respectable objective [4]. Online media is likewise improving the world's economy through setting out many new position open doors [5].

In recent times, cyberbullying has become a concerning trend that involves analyzing people's attitudes and opinions on social media platforms such as Face book, Twitter, and blogs. The

ISSN (Online): 2583-7052

International Journal of Advanced Trends in Engineering and Management (IJATEM)
pp. 132-143

Article Title: Detection of Cyber bullying on Social Media using Deep Learning Algorithms

primary objective of cyberbullying is to identify polarity, i.e., positive, negative, or neutral. Sarcasm is a specific type of cyberbullying that can reverse the polarity of a given text. Cyberbullying refers to sending misinformation to an individual or a community that causes heated debates among users. The growth of social media platforms like Face book, Instagram, and Twitter has increased cyberbullying, making it common among teens. This aggressive behaviour has a psychological and critical impact on the victim and can also be imitated by other members of the group. According to a new global data survey, 4,444 cyberbullying cases are on the rise every day. Many young people spend their time online sharing information on social networks.

Social networks allow people to communicate and share information with anyone at any time. There are over 3.444 billion social media users worldwide. According to the National Criminal Security Council (NCPC), cyberbullying is an online activity that involves hurting or intentionally embarrassing another person through a cell phone, video game application, or other means of sending or sending a text, photo, or video. Cyberbullying can occur anytime and anywhere, and it is possible to reach anyone with internet access. Texts, photos, or videos of cyberbullying can be published in unknown ways, and tracing their source can be difficult or impossible. Moreover, it is often not possible to delete these messages later. The most popular websites for bullying on the internet are Twitter, Instagram, Face book, YouTube, Snap chat, Skype, and a number of social networking sites. However, certain social networking sites like Facebook offer bullying prevention guidance, including a special section on how to report cyberbullying and prevent it from blocking users. Similarly, on Instagram, users can monitor or block individuals who share photos and videos that make them uncomfortable. Additionally, users can suggest improvements for the app and report violations to the community.

2. Literature Survey

The following survey helped us in finding the right set of sensors and modules for building our proposed model.

[1] Andrea Pereraa 2021 Cyber bullying has emerged as a pervasive and harmful phenomenon on social media platforms, posing serious threats to individuals' mental health and societal well-being. This paper presents a comprehensive study on the development of an accurate cyber bullying detection and prevention system tailored for social media environments. We begin by reviewing the existing literature on cyber bullying detection techniques, highlighting their limitations and shortcomings. Building upon this foundation, we propose a novel approach that leverages advanced natural language processing (NLP) and Deep methodologies. Our system combines convolutional neural networks (CNNs) and recurrent neural networks (RNNs) to analyze textual content, images, and videos posted on social media, thereby achieving a more holistic understanding of cyber bullying behaviours. To address the challenges associated with context and language nuances in social media conversations, we introduce a hybrid model that

incorporates word embeddings, attention mechanisms, and contextual analysis. This hybrid model improves the accuracy of cyber bullying detection by capturing subtle contextual cues and identifying evolving trends in online harassment.

[2] Jalal Omer Atoum 2020Cyberbullying is a growing concern in the digital age, with social media platforms providing fertile ground for harassment and harm. This paper presents an innovative approach to cyber bullying detection through sentiment analysis, utilizing natural language processing techniques to identify and mitigate online harassment. The research begins by outlining the prevalence and detrimental effects of cyber bullying on individuals and society. We highlight the importance of early detection and intervention in mitigating the psychological and emotional consequences faced by victims. Our proposed method employs sentiment analysis to assess the emotional tone of user-generated content on social media platforms. We explore the use of Deep learning algorithms to classify social media posts into categories such as positive, negative, or neutral sentiment. By training models on large datasets, we develop an effective framework for distinguishing cyber bullying content from harmless interactions. Leveraging sentiment analysis allows us to uncover subtle nuances in language and context, enabling more accurate cyber bullying detection. Moreover, we discuss the challenges associated with sentiment analysis in the context of cyber bullying, including the need to adapt to evolving online behaviours and account for cultural differences in language use. We present strategies to enhance the model's adaptability and scalability to different social media platforms and user demographics. To evaluate the effectiveness of our approach, we conducted extensive experiments on diverse social media datasets, measuring the system's precision, recall, and overall performance. The results demonstrate the promise of sentiment analysis as a viable tool for cyber bullying detection, with the potential to reduce false positives and improve intervention strategies. In conclusion, this paper offers a novel perspective on combating cyber bullying through sentiment analysis. By leveraging the power of natural language processing and Deep learning, we present an efficient and adaptable method for identifying cyber bullying behaviours on social media. This research contributes to the ongoing efforts to create safer and more respectful online communities, ultimately fostering a digital environment where users can engage without fear of harassment or harm.

3. Existing System

Multimedia algorithms, such as lower face detection accuracy for women and people of color, have highlighted the urgent need to develop strategies that are equally effective for various demographic groups. As a result, we assert that a significant research problem is assuring fairness in multimodal cyberbullying detectors (e.g., equal performance regardless of the victim’s gender). We suggest a fairness-aware fusion architecture that makes sure accuracy and justice are still crucial factors to take into account when integrating data from various modalities. The inputs from several modalities are integrated in this Bayesian fusion framework while taking into account the interdependencies between features and the various confidence

levels connected to each feature. Specifically, this framework assigns weights to different modalities not just based on accuracy but also their fairness. Results of applying the framework on a multimodal (visual) cyberbullying detection problem demonstrate the value of the proposed framework in ensuring both accuracy and fairness. In particular, this approach distributes weights to various modalities depending on both their fairness and accuracy. Results from using the framework to solve a multimodal (visual) cyberbullying detection problem show its value in assuring fairness and accuracy.

3.1 Block Diagram

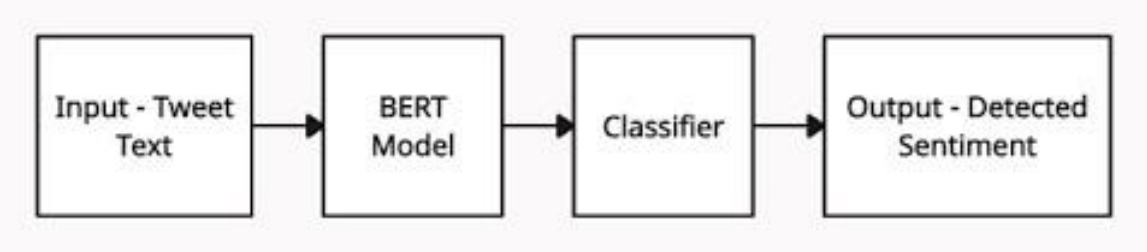

Figure 1: *Existing Block Diagram*

4. Proposed Work

Online apps for socializing are a place where we can express our thoughts and opinions and can even share our personal lives with another person [1]. We can access social media with the help of internet connection in our phones, laptops, PCs, tablets etc. The most well-known online media incorporates Facebook1, Twitter2, Instagram3, Tik-Tok4, etc. These days, web-based media is engaged with various areas like Coaching [2], Entrepreneurship [3], and furthermore for the respectable objective [4]. Online media is likewise improving the world's economy through setting out many new position open doors [5].

Albeit web-based media has a great deal of benefits, it likewise has a few downsides. Utilizing this media, malignant clients lead exploitative and deceitful demonstrations to offend and harm their notoriety. As of late, cyberbullying emerges as the significant web-based social apps issue.

Cyberbullying or digital badgering alludes to an electronic strategy for tormenting or provocation. Cyberbullying and digital badgering are otherwise called internet tormenting. With the boom and advancement of the social apps like Twitter, Face book, cyberbullying has become normal in the life, especially in the life of students.

ISSN (Online): 2583-7052

International Journal of Advanced Trends in Engineering and Management (IJATEM)
pp. 132-143

Article Title: Detection of Cyber bullying on Social Media using Deep Learning Algorithms

4.1 Block Diagram

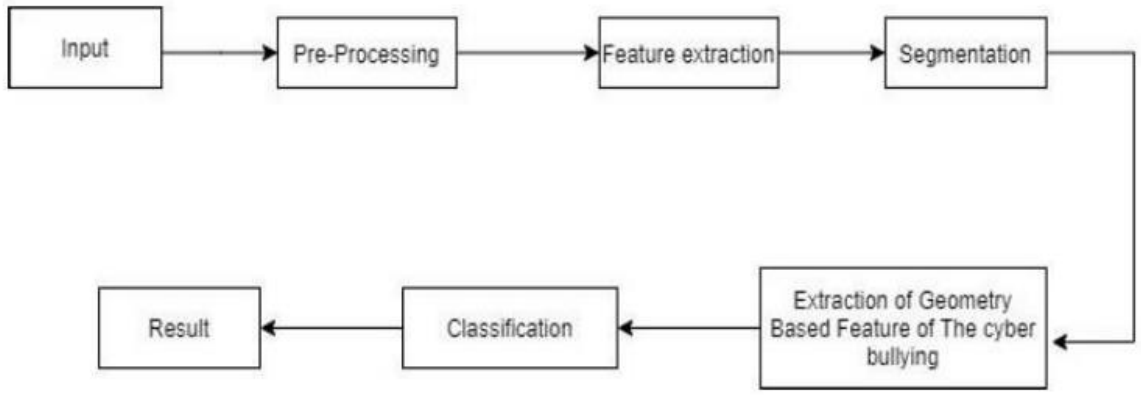

Figure 2: *Proposed Block Diagram*

The proposed method in this study represents an innovative approach to cyberbullying detection on social media platforms. Unlike previous research, which often focused solely on textual features, this method integrates three distinct types of features: textual, behavioral, and demographic. Textual features encompass the linguistic characteristics of messages or posts, including the presence of intimidating words or phrases commonly associated with cyberbullying. By analyzing the content of messages, the algorithm can identify patterns indicative of potential cyber bullying behaviour.

Behavioral features capture users' online actions and interactions, providing additional context for identifying cyberbullying instances. These features may include factors such as frequency of posting, engagement with specific content, or patterns of interaction with other users. Demographic features include information such as age, gender, and location extracted from user profiles or metadata associated with social media posts. By considering demographic factors, the algorithm can better understand the socio-cultural context in which cyberbullying occurs and tailor detection strategies accordingly.

By combining these multifaceted features, the proposed method aims to enhance the accuracy and effectiveness of cyberbullying detection algorithms. Deep learning algorithms, such as supervised classifiers or Deep models, can be trained on labelled datasets to recognize patterns and distinguish between normal interactions and instances of cyberbullying.

Furthermore, the proposed method addresses previous limitations in cyberbullying research, such as data unavailability and concerns about victim privacy, by leveraging comprehensive datasets and anonymization techniques to protect user identities. Overall, the proposed method represents a significant advancement in cyberbullying detection, offering a holistic approach that considers both textual content and contextual factors to more accurately identify and mitigate instances of online harassment and abuse.

ISSN (Online): 2583-7052

International Journal of Advanced Trends in Engineering and Management (IJATEM)
pp. 132-143

Article Title: Detection of Cyber bullying on Social Media using Deep Learning Algorithms

4.2 Implementation

The proposed method in this study represents an innovative approach to cyberbullying detection on social media platforms. Unlike previous research, which often focused solely on textual features, this method integrates three distinct types of features: textual, behavioral, and demographic. Textual features encompass the linguistic characteristics of messages or posts, including the presence of intimidating words or phrases commonly associated with cyberbullying. By analyzing the content of messages, the algorithm can identify patterns indicative of potential cyber bullying behavior. Behavioral features capture users' online actions and interactions, providing additional context for identifying cyberbullying instances. These features may include factors such as frequency of posting, engagement with specific content, or patterns of interaction with other users.

Demographic features include information such as age, gender, and location extracted from user profiles or metadata associated with social media posts. By consideringdemographic factors, the algorithm can better understand the socio-cultural context in which cyberbullying occurs and tailor detection strategies accordingly. By combining these multifaceted features, the proposed method aims to enhance the accuracy and effectiveness of cyberbullying detection algorithms. Deep learning algorithms, such as supervised classifiers or Deep models, can be trained on labeled datasets to recognize patterns and distinguish between normal interactions and instances of cyberbullying.

Furthermore, the proposed method addresses previous limitations in cyberbullying research, such as data unavailability and concerns about victim privacy, by leveraging comprehensive datasets and anonymization techniques to protect user identities. Overall, the proposed method represents a significant advancement in cyberbullying detection, offering a holistic approach that considers both textual content and contextual factors to more accurately identify and mitigate instances of online harassment and abuse.

5. Software and Hardware Description:

5.1 Requirements

Hardware

- Hard Disk-1GB
- RAM-4GB
- LAPTOP
- Windows-11 OS

Software

- Python Language
- OPEN CV
- ANACONDA
- Pytorch

ISSN (Online): 2583-7052

International Journal of Advanced Trends in Engineering and Management (IJATEM)
pp. 132-143

Article Title: **Detection of Cyber bullying on Social Media using Deep Learning Algorithms**

Python

Python is a high-level, interpreted, interactive and object-oriented scripting language. Python is designedto be highly readable. It uses English keywords frequently where as other languages use punctuation, and it has fewer syntactical constructions than other languages.

Python is Interpreted − Python is processed at runtime by the interpreter. You do not need to compile your program before executing it. This is similar to PERL and PHP.

Python is Interactive − you can actually sit at a Python prompt and interact with the interpreter directly to write your programs.

Python is Object-Oriented − Python supports Object-Oriented style or technique of programming that encapsulates code within objects.

Python is a Beginner's Language − Python is a great language for the beginner-level programmers' andsupports the development of a wide range of applications from simple text processing to WWW browsers to games.

6. Simulation Results

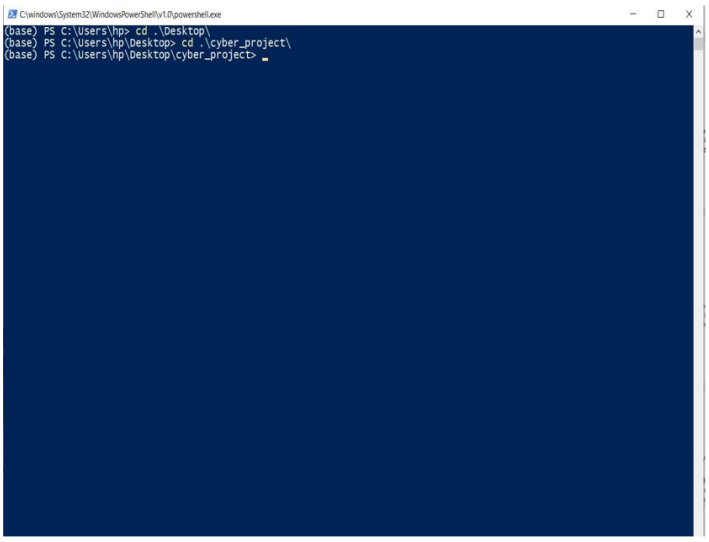

Figure 3: *Command Prompt of File*

- Here it is a command prompt where we have to be connect with the server to the client file.
- Type CD Desktop to get the file location.
- Type the CD Cyber project to extract the file.
- Type Python Server to make connection with local server.
- Type Python Client GUI to get connects both server and client command prompt.

ISSN (Online): 2583-7052

International Journal of Advanced Trends in Engineering and Management (IJATEM)
pp. 132-143

Article Title: Detection of Cyber bullying on Social Media using Deep Learning Algorithms

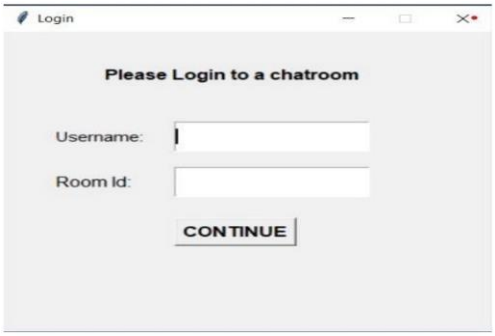

Figure 4: *Login Page*

- Enter the prompt after entering the 3 commands it will pop an application as login page.
- Enter the Username and Room Id.

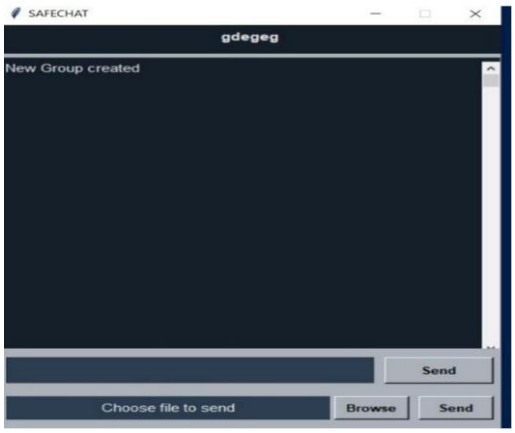

Figure 5: *Safe Chat Page*

- Here an application pop out by the name safe chat it is a comment box where we have provide the input.

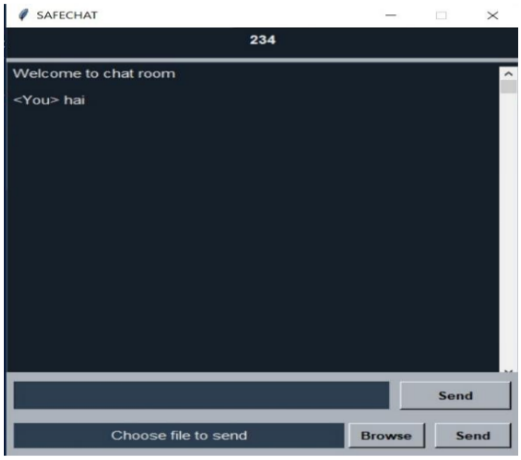

Figure 6: *Non Bullying Chat*

ISSN (Online): 2583-7052

International Journal of Advanced Trends in Engineering and Management (IJATEM)
pp. 132-143

Article Title: Detection of Cyber bullying on Social Media using Deep Learning Algorithms

- Here we providing the comment as Hay it will detect and predict whether the given comment is bullying or not and it sends an acknowledgement to the client command prompt.

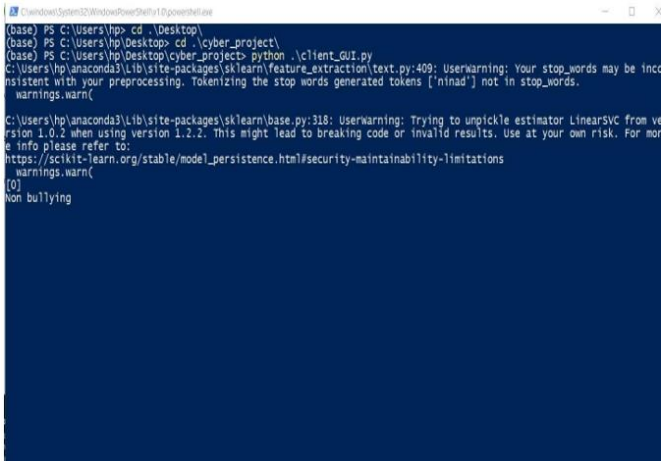

Figure 7: *Non Bullying Command Prompt*

- From the above command prompt we can observe that the given input tweet is a Non-Bullying chat.

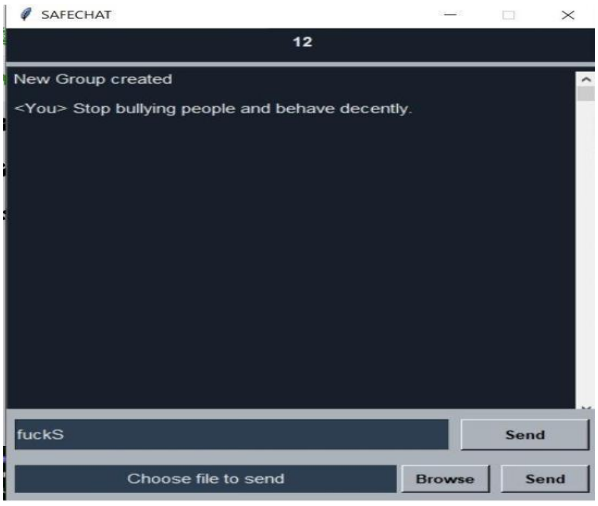

Figure 8: *Bullying Chat*

- Here we have provided the bullying chat so it is displaying in the comment box as stop bullying the people behave decently.

ISSN (Online): 2583-7052

International Journal of Advanced Trends in Engineering and Management (IJATEM)
pp. 132-143

Article Title: **Detection of Cyber bullying on Social Media using Deep Learning Algorithms**

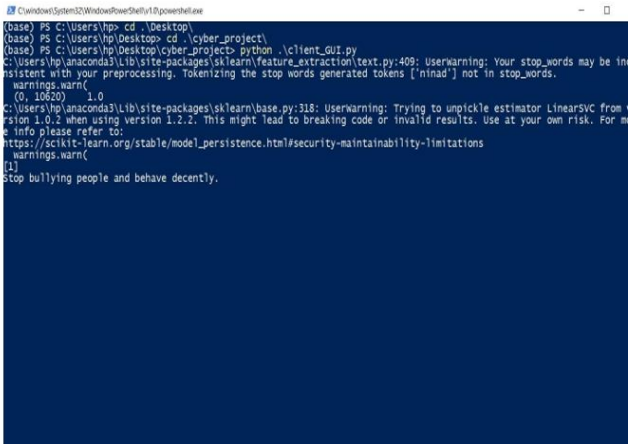

Figure 9: *Bullying command prompt*

- From above command we can observe that the post is not uploaded in the comment box and it is sending acknowledgement to client as stop bullying people behave decently**.**

6.1 Graphical Representation

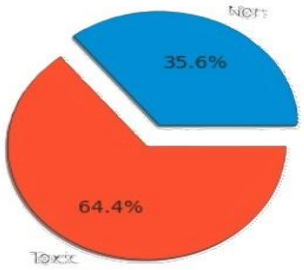

Figure 10: *Toxic and Non-Toxic*

- The above graph represents the given comment is toxic and Non-Toxic in the form of percentage.

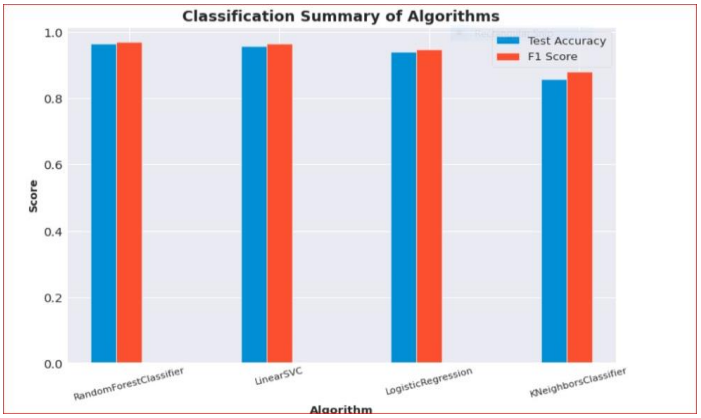

Figure 11: *Classification Summary of Algorithms*

- The above bar graph represents the classification summary of Algorithms in the form of Test Accuracy and F1 score.

7. Conclusion

The development and implementation of deep learning-based cyber bullying detection systems hold significant promise in addressing the pervasive issue of online harassment and abuse. By leveraging advanced algorithms and data-driven approaches, these systems can accurately identify instances of cyber bullying on social media platforms, thereby facilitating timely intervention and support for affected individuals. The advantages of Deep learning-based detection, including efficiency, accuracy, scalability, and customization, make them valuable tools for creating safer online environments for users of all ages. Furthermore, the diverse applications of Deep learning-based cyber bullying detection span across various sectors, including social media platforms, educational institutions, parental control tools, corporate environments, law enforcement, community support services, and research and policy development. By deploying these systems in appropriate contexts, stakeholders can proactively address cyberbullying, protect vulnerable populations, and foster a culture of respect and empathy in digital spaces

8. Future Scope

The Further research is needed to improve the robustness and generalization ability of Deep learning models, particularly in handling diverse linguistic styles, cultural nuances, and evolving patterns of cyber bullying behaviour. Future systems could explore the integration of multimodal data sources, including text, images, and audio, to capture a more comprehensive understanding of cyberbullying incidents and improve detection accuracy

References

1. Y. Bengio, A. Courville, and P. Vincent, “Representation learning: A review and new perspectives,” Pattern Analysis and DeepIntelligence, IEEE Transactions on, vol. 35, no. 8, pp. 1798–1828, 2013

2. A. M. Kaplan and M. Heinlein, “Users of the world, unite! The challenges and opportunities of social media,” Business horizons, vol. 53, no. 1, pp. 59–68, 2010.

3. R. M. Kowalski, G. W. Giumetti, A. N. Schroeder, and M. R.Lattanner, “Bullying in the digital age: A critical review and misanalysis of cyber bullying research among youth.” 2014

4. B. K. Biggs, J. M. Nelson, and M. L. Sampilo, “Peer relations in the anxiety– depression link: Test of a mediation model,” Anxiety, Stress, Coping, vol. 23, no. 4, pp. 431–447, 2010.

5. K. Dinakar, B. Jones, C.Havasi, H. Lieberman, and R. Picard. ”Common sense reasoning for detect ion, prevent ion, and mitigate ion of cyber bullying.” ACM Transact ions on Interactive Intelligent Systems (TiiS) 2, no. 3, 2012, p. 18.

ISSN (Online): 2583-7052

International Journal of Advanced Trends in Engineering and Management (IJATEM)
pp. 132-143

Article Title: Detection of Cyber bullying on Social Media using Deep Learning Algorithms

6. V. Nahar, S. Unankard, X. Li, and C. Pang. ”Sentiment analysis for effective detect ion of cyber bullying.” In Asia-Pacific Web Conference, Springer, Berlin, Heidelberg, 2012, pp. 767-774.

7. V. Nahar, X. Li, C. Pang, and Y. Zhang. ”Cyberbullying detect ion based on text-st ream classification.” In The 11th Australasian Data Mining Conference (AusDM 2013), 2013.

8. M. Dadvar, D. Trieschnigg, R. Ordelman, and F. de Jong. ”Improving cyberbullying detect ion with user context.” In European Conference on Information Retrieval, Springer, Berlin, Heidelberg, 2013, pp. 693-696.

9. V. Nahar, S. Al-Maskari, X.Li, and C.Pang. “Semi-supervised learning for cyberbullying detect ion in social networks.” In Australasian Database Conference, Springer, Cham, 2014, pp. 160- 171.

10. V. Nahar, X. Li, H. L. Zhang, and C. Pang. “Detecting cyberbullying in social networks using multi-agent system.” Web Intelligence and Agent Systems. An International Journal 12, no. 4, 2014, pp. 375-388.