JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue

JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)

An International Scholarly Open Access, Peer-reviewed, Refereed Journal

"ENHANCING ONLINE SAFETY: DEVELOPING AN AUTOMATED SYSTEM TO DETECT CYBERBULLYING ON SOCIAL MEDIA"

Sanika V. Pande[1], Trupti B. Bhagat[2], Shripad J. Khorgade[3], Vaishanavi D. Rajgure[4] , Samiksha M. Gopal[5] , Chanchal A. Kshirsagar[6]

[1,2,3,4,5]Student(UG) Computer Engineering Department Jagadambha College of Engg. & Tech. Yavatmal

[6]Assistant Professor Computer Engineering Department Jagadambha College of Engg. & Tech. Yavatmal

Abstract : Contemporary youth, commonly referred to as "digital natives," have come of age in a world immersed in advanced technologies. They are well-versed in real-time communication and exhibit no constraints when it comes to connecting with others or online communities. Nonetheless, the proliferation of social media platforms has rendered them more susceptible to cyberbullying, a form of harassment facilitated by technology. While this issue has persisted for some time, our comprehension of the profound impact it has on young individuals has only recently deepened. Leveraging machine learning techniques, we can discern patterns in the language employed by both aggressors and their targets, enabling us to establish automated protocols for identifying cyberbullying content. Cyberbullying, an online manifestation of harassment, presents a grave concern with far-reaching ramifications. It manifests in diverse ways, frequently through written communications on social networks. Given the staggering 1.96 billion individuals active on social media, its prevalence is undeniable. As we transition into the next decade, we encounter substantial challenges in addressing online conduct. Incidents of taunting, harassment, and emotional harm inflicted online are on the rise. The automatic detection of these occurrences necessitates the deployment of sophisticated computer systems.

Keywords: cyber threat, bullying, victim, mitigation, machine learning.

Introduction :

The internet has undeniably become an integral part of our lives, offering immense opportunities, but also harboring a dark side, particularly evident in the form of cyberbullying. Cyberbullying is a pressing concern due to its online nature, facilitated by technology that has made it distressingly effortless for individuals to inflict harm on one another. Social media platforms like Facebook, Twitter, Instagram, have become fertile grounds for cyberbullying to thrive, necessitating comprehensive examination and comprehension of this complex issue. One major challenge in addressing cyberbullying is the absence of specific laws in many countries, further complicating efforts to combat this menace. Cyberbullying encompasses various forms of online communication intended to intimidate or threaten others, including sexual harassment, creating hostile environments, seeking revenge, or engaging in retaliatory actions. One of its insidious characteristics is the anonymity it affords perpetrators, rendering it difficult to identify and rectify. The ramifications of cyberbullying are severe, potentially leading to self-harming thoughts and actions. Detecting and preventing cyberbullying is imperative to provide support to those who are suffering. Unlike traditional bullying, cyberbullying unfolds in the virtual realm, rapidly disseminating and often operating under the veil of anonymity. This research concentrates on the identification of cyberbullying through text-based communication, which represents the most prevalent form of this phenomenon. To achieve this, Convolutional Neural Networks (CNNs), a subtype of artificial intelligence, are deployed to analyze textual data. The ultimate objective is to develop systems capable of recognizing and thwarting instances of cyberbullying, with a primary focus on safeguarding the well-being of young individuals. This endeavor holds profound significance because cyberbullying often transpires on social networks that are challenging to monitor comprehensively, causing substantial harm to victims' mental and emotional health. It can instigate persistent anxiety, stress, sadness, self-doubt, and feelings of inadequacy. Furthermore, it may lead to social isolation, self-harm contemplation, adverse academic or professional performance, and even physical ailments such as headaches and sleep disturbances. Some individuals may abstain from activities they once enjoyed due to fear of encountering cyberbullying. In the gravest scenarios, it can result in traumatic memories and nightmares, akin to those experienced by individuals subjected to distressing situations. In conclusion, cyberbullying represents a grave issue with far-reaching implications for both mental and physical well-being, necessitating urgent intervention and support for those affected.

LITERATURE SURVEY :

[1] Vijay B, Jui T, Pooja G, Pallavi V., "Detection of Cyberbullying Using Deep Neural Network" examines :

In this study, a different method was used to find cyberbullying. They used something called as "convolutional neural network," which is like a smart computer system. It looks at things through many layers and can tell what something is more accurately. This is smarter than the old way of figuring things out.

[2] Sabina T, Lise G, Yunfei C, Yi Z. "A Socio-linguistic Model for Cyberbullying Detection" examine :

AI models face challenges when dealing with social media data, which often includes short messages with mistakes and slang. Detecting cyberbullying is even harder because we rely on other people to label the data for training. To address these issues, two types of models were used: space-inspired semantic models and socio-phonetic model. The space-inspired models make it easier by reducing the number of things the model needs to learn and by using connections between words and messages. The socio-phonetic model can figure out relationships from limited social media data while spotting cyberbullying. As far as we know, it's the first model to understand bullying content, categories, user roles, and relationships all together. By doing this, it can learn from social dynamics and significantly improve cyberbullying detection and understanding who is involved.

[3] Batoul H, Maroun C, Fadi Y., "Cyberbullying Detection: A Survey on Multilingual Techniques" examine :

This project introduced a system that can detect cyberbullying in different languages. It uses techniques from machine learning (ML) and natural language processing (NLP) to do this. The system is good at finding cyberbullying on social media sites like Facebook and Twitter. It can work with content in Arabic, English, or a mix of both, including Arabist or Arabize texts.

[4] Lin L., Linlong X., Nanzhi W., Guocai Y. "Text classification method based on convolution neural network" examine :

This research focuses on a technique that uses CNN, a type of computer system, for organizing content without needing to remove specific parts beforehand. Here is how it works: First, it prepares messages using something called "Word2vec" to create word patterns based on Chinese characteristics. Then, it assigns importance to these word patterns using a measure called "TFIDF." Finally, it uses CNN to extract higher-level patterns and improve the system's ability to organize information. To prevent the system from becoming too specialized, a method called "Dropout" is used. This technique is tested in different situations, like varying text lengths and emphasis, and is compared to other ways of organizing content.

[5] Lu C., Jundong L., Yasin N. S., Deborah H., Huan L., " XBully: Cyberbullying Detection within a Multi-Modal Context" examine :

This study looked at a new problem: finding cyberbullying in a mix of different types of online content. To handle this, they came up with a new system called "X Bully" that learns from the connections between different types of content. X Bully first identifies key sources of information to deal with different types of features. Then, it puts everything into the same space to understand how different parts relate to each other. They tested this on real-world data and found it works well. In the future, they want to understand more about how different aspects of cyberbullying work, like the roles of different people involved, such as victims and bullies. They also believe that solving this problem will require experts from both computer science and psychology working together.

[6] Mohammed A. Kasturi Dewi V., Sri Devi R., "Cybercrime detection in online communications" examine :

This study created a model to find cyberbullying on Twitter. They used different aspects of tweets, like how they are organized, who wrote them, and what they say, to build a smart computer program (AI) that can tell if a tweet is cyberbullying or not. They tested four different classifiers to see which one works best: NB, Lib SVM, random forest, and KNN. They also used three methods to figure out which features are most important: a c2 test, information gain, and Pearson correlation. This helped them identify the key elements that help detect cyberbullying in tweets.

[7] Monirah A., Mourad Y., "Optimized Twitter Cyberbullying Detection based on Deep Learning" examine :

In this research, they looked at how people currently find cyberbullying on Twitter and suggested a new way using deep learning. They created a method called OCDD, which uses data labeled by humans to teach the computer. They turned each word into something called "word vectors" using a technique called GloVe. Then, they used a special computer program called a convolutional neural network (CNN) to figure out if a tweet is cyberbullying or not. OCDD makes it easier to find cyberbullying because it does not need traditional methods of picking out important words. Instead, it uses word patterns and CNN, who is smarter than the usual methods. While CNN has been successful in other areas, this research shows it can also work well for finding cyberbullying.

[8] Xiang Z, Jonathan T, Nishant V, Elizabeth W., "Cyberbullying Detection with a Pronunciation Based Convolutional Neural Network" examine :

This study introduced a new way to use speech patterns to find cyberbullying using a special computer program called a convolutional neural network (CNN). They compared this method with two other CNN models and different classifiers using two sets of data, each with varying levels of mistakes and differences in the number of cyberbullying cases. This new method worked really well on the datasets they used. Additionally, they tried three strategies to handle situations where there were not many cyberbullying cases compared to non-cyberbullying ones. The results showed that the PCNN with cost function adjustment was a good solution for this problem.

PROBLEM STATEMENT & OBJECTIVES:

Cyberbullying is an extensive and grave concern within the digital sphere, manifesting through various channels such as online chats, text messages, and emails. Notably, social media giants like Facebook, Twitter and Instagram have emerged as fertile grounds for the proliferation of cyberbullying incidents. This surreptitious and harmful online behavior inflicts substantial harm, with young individuals bearing the brunt, enduring emotional distress, anxiety, and, alarmingly, harboring self-harming thoughts. Unlike conventional bullying, which typically occurs in school environments and falls under the scrutiny of educators and parents, cyberbullying thrives in the digital realm, evading the watchful eyes of responsible adults. At the core of this issue is the formidable challenge of conceiving and implementing an intelligent system endowed with the capability to autonomously identify instances of cyberbullying on social media platforms. The overarching objective is to foster a digital landscape that is safer, more compassionate, and respectful for all users.

The primary goals of this project include gaining a deeper understanding of cyberbullying by examining its different forms and prevalence on social media, as well as reviewing existing research on effective bullying detection methods for adaptation. We aspire to develop an automated system employing advanced technologies such as machine learning and natural language processing to swiftly identify instances of cyberbullying on social media platforms. Additionally, we aim to integrate real-time reporting features to enable prompt responses to cyberbullying incidents. Ensuring the user-friendliness and seamless integration of the system into social media platforms is another crucial objective. We will consistently evaluate and enhance the system's performance based on real-world data and user input. In parallel with the technical aspects, we will actively raise awareness about cyberbullying and advocate for responsible online behavior to cultivate a more respectful and compassionate online community, with a particular focus on safeguarding young individuals who are vulnerable to cyberbullying.

A. EXISTING SYSTEM :

Many recent studies have used common AI models, and most of the models they created work for just one social network. Deep learning models have also been used to find instances of cyberbullying, suggesting they can do a better job than traditional models in detecting such incidents. However, even with these advanced models, there is still challenge in effectively categorizing the bullying behavior. To tackle cyberbullying and prevent its harmful consequences online, machine learning and language processing techniques are being used. But, it's important to note that the parameters involved in these methods can be quite complex and challenging to understand, especially when dealing with confusing or subtle forms of bullying.

B. PROPOSED SYSTEM :

In our system, we use something called CNN, which stands for Convolutional Neural Network. This CNN has multiple layers, and it works by analyzing information step by step through these layers. This process is kind of like how the central nervous system in mammals works. Deep learning, which is what this kind of network is part of, involves many layers of learning on its own. So, in simple terms, our system uses CNN to carefully analyze data in layers, inspired by how mammals' nervous system work, and it's a part of deep learning, which can learn a lot by itself.

a. Data pre-processing :

Data pre-processing is like cleaning up data before using it for any task, and it's super important. It's the first thing you do to get data ready for teaching a computer program. Imagine you have a messy sentence like "You look so ugly and fat, change the style." Pre-processing makes it looks like this: "look ugly fat changestyle." It gets rid of unimportant words like "as," "what," "who," and special symbols like @, (), [], ?/;, which the computer doesn't need to learn. Also, it breaks the text into sentences and makes sure each sentence has the same number of words by adding a common word. This helps keep everything neat and tidy. The computer likes data in a certain format, so pre-processing turns the text into lowercase and changes it into a special format called a vector, which the computer can work with. So, data pre-processing is all about cleaning up and organizing data to make it usable for the computer.

b. CNN Model Layers :

The core of this entire process hinges on the utilization of Convolutional Neural Network (CNN) layers, which play a vital role in data processing. Visualize these layers as fundamental building blocks for your computer program. Among the key layers employed in this model, we have the Sequential Layer, which serves as a foundational element for a framework known as Keras. Keras, in turn, is a powerful tool for constructing neural networks, which can be likened to computer systems that possess learning capabilities. The Sequential Model, at its essence, is quite straightforward. It's akin to stacking multiple layers atop one another, and it's referred to as "dense" because all the nodes within each layer establish connections with nodes in other layers. Envision it as an intricate web of interconnections between these nodes. In this context, there's a crucial element known as a perceptron. A perceptron is a singular algorithm that takes a multitude of values as input and yields either a "yes" (1) or "no" (0) as its output. However, there's a caveat: perceptrons are effective for handling small datasets but tend to falter when confronted with substantial amounts of data. They struggle to glean insights from such extensive information. Furthermore, their binary output of 0 or 1 is not always suitable for all applications. To address these limitations, we introduce a fundamental concept: the activation function. This function acts as a smoothing agent, enhancing the versatility of the neural network. Various types of activation functions exist, including sigmoid and ReLu (Rectified Linear Unit). The sigmoid function yields continuous values rather than a binary outcome, which proves invaluable for performing calculations within the neural network. ReLu, on the other hand, functions as another activation function, operating on a straightforward principle: if the input is negative, it outputs 0; if it's positive, it progressively amplifies. This avoids abrupt transitions in the data. Within this network, we take input text and transform it into a sequence of numerical representations corresponding to words. To accomplish this, we employ a tool called NLTK (Natural Language Toolkit) to segment the text into sentences and subsequently into individual words. This preprocessing step allows the computer to comprehend and effectively manipulate the textual information. In essence, the CNN model layers are pivotal in the data processing pipeline, with the Sequential Layer and activation functions facilitating the creation of a versatile and effective neural network for text analysis.

c. Model Prediction :

Once we've defined our model, the next step is to compile it. This is like getting it ready to do the actual work using a tool called Keras. Keras can run on different backends like Theano or TensorFlow. Compiling involves setting up a few important things:

Optimizers: These are like helpers that adjust the model's weights while it's learning. They help our model get better.

Loss Function: This tells the model how well or poorly it's doing. It's like a scorecard that guides the learning process.

Metrics: These are measurements that help us understand how good our model is performing.

Once our model is compiled, we can start training it with the "fit()" function. There are some important parameters we use:

Epochs: This is how many times our model gets to learn from the training data. It's like going through the training data multiple times to get better.

Batch Size: This is how many training examples our model looks at before making adjustments. It's like learning in small groups instead of one at a time.

Validation Data: This is data we use to test how well our model is learning. We compare the results to see if it's doing a good job. This process repeats until our model gets really good or reaches the best it can be. So, compiling and training are essential steps in getting our model ready to do its job effectively.

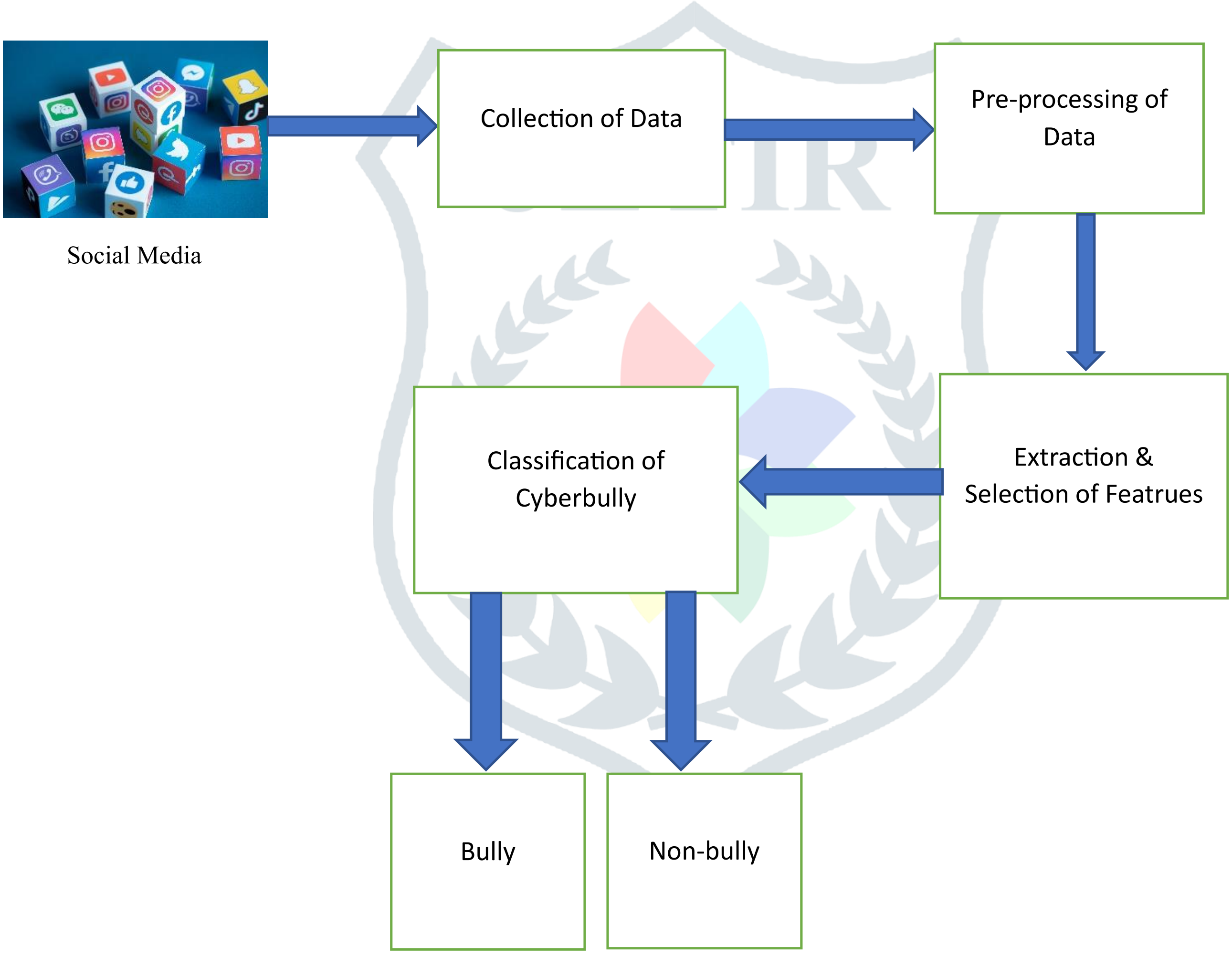

Figure 1

CONCLUSION :

The technology revolution has undoubtedly improved our lives, but it has also created opportunities for criminals to commit crimes, especially on the internet. Internet crimes are particularly concerning because victims can be continuously targeted, and it's often hard to escape from these situations. Cyberbullying is a particularly serious internet crime, and research has shown its devastating impact on victims. In this paper, a new idea is presented to identify cyberbullying comments in tweets. The system uses a highly accurate method involving Convolutional Neural Networks (CNN) implemented using Keras, a machine learning tool. This approach aims to provide precise results in detecting cyberbullying. The proposed system can be valuable not only for government agencies but also for parents, guardians, schools, policymakers, and law enforcement. By using this system, users can be better protected from the harsh consequences of cyberbullying. Since online bullying is an ongoing issue, it's crucial to continually update and improve the methods used to combat it. Our proposed approach can be a valuable tool for managing such crises and could potentially offer round-the-clock support. Ultimately, it has the potential to prevent cyberbullying crises from happening in the first place.

REFERENCES :

[1] Vijay B, Jui T, Pooja G, Pallavi V., ”Detection of Cyberbullying Using Deep Neural Network”, in 5th International Conference on Advanced Computing & Communication Systems (ICACCS), pp.604-607, 2019

[2] Sabina T, Lise G, Yunfei C, Yi Z. ”A Socio-linguistic Model for Cyberbullying Detection”, in the IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), 2018

[3] Batoul H, Maroun C, Fadi Y., ”Cyberbullying Detection: A Survey on Multilingual Techniques” in European Modelling Symposium (EMS), pp. 165–171, 2016

[4] Lin L., Linlong X., Nanzhi W., Guocai Y. ”Text classification method based on convolution neural network”, in 3rd IEEE International Con-ference on Computer and Communications (ICCC), pp . 1985-1989, 2017

[5] Lu C., Jundong L., Yasin N. S., Deborah H., Huan L.,” XBully: Cyberbullying Detection within a Multi-Modal Context”, in Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining, pp. 339-347, 2019

[6] Mohammed A. Kasturi Dewi V., Sri Devi R., ”Cybercrime detection in online communications”, in the Computers in Human Behavior, 2016

[7] Monirah A., Mourad Y., ”Optimized Twitter Cyberbullying Detection based on Deep Learning” in 21st Saudi Computer Society National Computer Conference (NCC), 2018

[8] Xiang Z, Jonathan T, Nishant V, Elizabeth W., ”Cyberbullying Detection with a Pronunciation Based Convolutional Neural Network”, in 15th IEEE International Conference on Machine Learning and Applications (ICMLA), pp. 740-745, 2016