International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2022): 7.942

Cyberbullying using Web Based Support Vector Machine using Recurrent Method

Dr. S. Brindha

Assistant Professor, Department of Computer Applications, SRMIST Faculty of Science and Humanities, SRM Institute of Science and Technology, SRM University, Kattankulathur, Chengalpattu, Chennai, India
Email: *brindhas1[at]srmist.edu.in*

Abstract: ***Cyberbullying is a growing concern in the digital age, necessitating effective detection and prevention measures. This research focuses on addressing cyberbullying through text classification using Natural Language Processing (NLP) algorithm - based techniques. The study involves the compilation of a diverse dataset containing social media posts, messages, and comments, which are labelled as either cyberbullying or non - cyberbullying content. The text data is pre - processed to handle noise, remove stop words, and tokenize the text for NLP analysis. Experimental results demonstrate the efficacy of NLP algorithm - based techniques in cyberbullying text classification, outperforming traditional rule - based methods. The system's ability to identify harmful content aids in early detection and intervention, promoting safer online environments. Experimental results demonstrate the efficacy of NLP algorithm - based techniques in cyberbullying text classification, outperforming traditional rule - based methods. The system's ability to identify harmful content aids in early detection and intervention, promoting safer online environments.***

Keywords: Cyberbullying, Text Classification, Natural Language Processing, NLP Algorithms, Naive Bayes, Support Vector Machines, Recurrent Neural Networks, Social Media, Online Safety

1. Introduction

In recent years the proliferation of social media platforms has transformed the landscape of human interaction, enabling unprecedented connectivity and communication across the globe. However, along the myriad benefits of this digital revolution comes the alarming rise of cyber bullying, a pervasive and harmful phenomenon that can have a devastating effects on individuals mental health and well - being.

In this pressing societal issue is to develop a robust cyber bullying text classification system using Recurrent Neural Networks (RNN). Leveraging the power of deep learning particularly RNNs, we seek to create a model capable of accurately identifying the instances of cyber bullying in textural data extracted from various online sources like as social media posts, comments, and messages. The endeavour to contribute the on - going efforts to combat cyber bullying by providing a scalable and efficient solution for automated text classification. By accurately identifying and flagging instances of cyber bullying in online discourse our system aims to empower platform administrators, moderators and utilizers to take proactive measures to mitigate the harmful effects of online harassment and promote a safer and more inclusive digital environment.

2. Related Works

Cyber bullying has emerged as a significant societal concern in the digital age, which studies highlighting its detrimental effects on individuals mental health well - being, and social interactions. As such researchers have increasingly turned to computational approaches particular natural language processing (NLP) and machine learning techniques to develop automated systems for detecting and classifying instances of cyber bullying in online text data.

For Traditional machine learning approaches its early efforts in cyber bullying detection often relied on traditional machine learning algorithms such as support vector machines, Naïve Bayes Classifiers and logistic regression. For instance the applied SVMs to classify cyber bullying instances in Twitter data [2], achieving promising results in terms of accuracy and precision.

For deep learning techniques have gained transaction for their ability to automatically learn intricate patterns and representations from raw text data. Recurrent Neural Networks (RNNs) have demonstrated remarkable performance in sequence modelling tasks, making them well suited for analysing textual data with temporal dependencies [5].

To further enhance the performance of RNN - based models attention mechanisms have been incorporated to selectively focus on relevant parts of the input sequence. Mustafa Ali Bamboat et. al [3] proposed an attention based RNN model for sentiment analysis demonstrating its capability to capture contextually important information and to improve classification accuracy.

In addition to transfer learning the model architectures transfer learning techniques have been explored to address the challenge of limited labled data in cyberbullying detection tasks. S. Asadianfam, H. Kolivand et. al [8] introduced Universal Language Model Fine - tuning (ULMFiT), a transfer learning approach that pretrains a language model on a large corpus of text data and fine tunes it on a specific task with limited labelled data. This approach has promising results in various NLP tasks and could be adapted to cyber bullying text classification.

Challenges of Cyberbullying:
Despite the progress in cyber bullying text classification, several challenges remain, including the detection of

nuanced forms of cyber bullying, handling multilingual text data and addressing biases in training data.

3. Proposed Work

Support vector machines have met with significant success in numerous real - world learning tasks. However, like most machine learning algorithms, they are generally applied using a randomly selected training set classified in advance. In many settings, we also have the option of using pool - based active learning. Instead of using a randomly selected training set, the learner has access to a pool of unlabeled instances and can request the labels for some number of them. We introduce a new algorithm for performing active learning with support vector machines, i. e., an algorithm for choosing which instances to request next.

1) Software Requirements:
Operating System: Windows 10 or later
Tool: Anaconda with Jupyter Notebook

2) Hardware requirements:
Processor: Intel i3
Hard disk: minimum 80 GB
RAM: minimum 2 GB

The Anaconda® distribution comes with a desktop graphical user interface (GUI) called Anaconda Navigator, which lets you manage conda packages, environments, and channels without the need for command - line interfaces. Navigator has the ability to look up packages locally in an Anaconda Repository or on Anaconda. org. Anaconda. Now, Anaconda is also a fantastic choice if your primary focus is data science. Continuum Analytics is the creator of Anaconda, a Python installation that has several helpful data science - related Python packages preinstalled [7]. With the goal of making package management and deployment easier, Anaconda is a distribution of the Python and R programming languages for scientific computing (data science, machine learning applications, large - scale data processing, predictive analytics etc.,

Stemming
Techniques for stemming words are employed to determine their root or stem. Words are stemmed, including a significant amount of language - specific linguistic information in the process. The idea behind stemming is that words can be confused by employing stems since words with the same stem or word root typically represent the same or very similar topics in literature. 'USE' is the root word for all of the following words: user, users, used, and using. The most widely used algorithm in English, the Porter Stemmer algorithm, is employed in this study. When two words with distinct stems are stemmed to the same root, this is known as over - stemming. Another term for this is a false positive. When two words that ought to stem to the same root do not, this is known as under - stemming. Another term for this is a false negative.

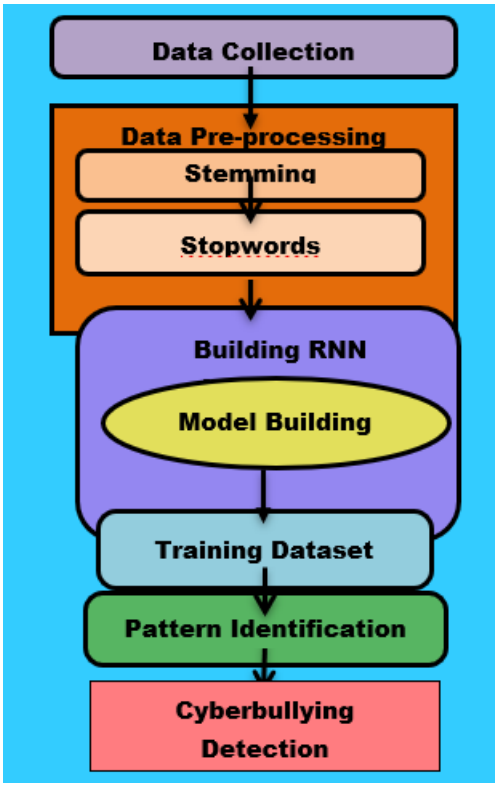

Figure 3.1: Enhanced Method using WBSVM Pattern Classification Method

3.2 Pattern Discovering

Recommendation systems built and generated suggestions based on information gleaned from user behavior and attributes by employing data mining techniques (association rules, sequential pattern finding, grouping, and classification). Recommender systems, put simply, attempt to infer the user's mental model, pinpoint his nearest and best interests, and then suggest his choices to him or her. The CBR approach is renowned for simulating how people would behave when faced with novel challenges. As a result, knowledge obtained from resolving past issues is utilized to inform the solution of current issues.

4. Results and Discussions

One kind of artificial neural network intended for processing sequential input is the Recurrent Neural Network (RNN). RNNs, in contrast to conventional feed forward neural networks, include connections that loop back on themselves. This feature enables them to keep track of information from earlier time steps in the sequence and retain a hidden state. RNNs are ideally suited for sequence - related applications including time series prediction, audio recognition, and natural language processing because of their looping mechanism. Here is a thorough breakdown of an RNN's architecture and essential parts:

Basic Structure:
An RNN is made up of several linked layers. It accepts an input vector (or sequence) and outputs an output vector (or sequence) at each time step t. What makes an RNN unique is its hidden state, represented by the letter "h. " The network's memory is represented by this concealed state.

Input and Output:
The RNN receives an input vector, or element x (t), at each time step t. A word in a phrase, a single element of a sequence, a pixel in a picture, etc. can all be considered this input. At every time step, the RNN generates an output vector, or element y (t). The output may be utilized for a number of purposes, including categorizing a sequence as a whole or forecasting the element that will come next.

Hidden State:
A vector called the hidden state h (t) is used to store data from earlier time steps. It functions as the network's memory. With the current input x (t) and the prior hidden state h (t - 1) as inputs, the hidden state is calculated at each time step.

Output Computation:
The current hidden state or a combination of the input and the hidden state at that time step can be used to calculate the output at each time step.

Back propagation Through Time (BPTT):
Back propagation via Time is a type of back propagation that is used in RNN training. Though it takes the sequential structure of the data into consideration, it is comparable to conventional backpropagation. To update the network's weights, the gradients are calculated at every time step and added up throughout the course of the sequence.

Future Cyberbulleying
Subsequent investigations may examine innovative model structures, methods for augmenting data, and cross - linguistic transfer learning strategies to surmount these obstacles and create more resilient and comprehensive cyberbullying identification systems.

References

[1] Dr. S. Brindha, Dr. S. Sukumarn, Relevance Pattern Discovery for Text Classification Using Taxonomy Methods" in International Journal for Science and Advance Research in Technology (IJSART) " Volume 4 Issue 11 –November 2018 ISSN [online]: 2395 - 1052.
[2] Dr. S. Brindha, Dr. S. Sukumaran, An Analysis on Big Data Interrogation Explore Problems and Tools, International Journal of Novel Research and Development (IJNRD), ISSN: 2456 - 4184, Volume 5, Issue 6, June 2020.
[3] Mustafa Ali Bamboat, Ghulam Sarfaraz Khan, Naadiya Mirbahar, Sheeba Memon, "Web Content Mining Techniques for Structured Data: A Review" (SJHSE) Sindh Journal of Headways in Software Engineering.2022; 1 (1).
[4] Richlin Selina Jebakumari A. Nancy Jasmine Goldena. A Survey on Web Content Mining Methods and Applications for Perfect Catch Responses. International Research Journal of Engineering and Technology (IRJET).2019; 06 (01): 407 - 412. e - ISSN: 2395 - 0056 p - ISSN: 2395 - 0072.
[5] Sharma PS, Yadav D, Thakur RN. Web Page Ranking Using Web Mining Techniques: A Comprehensive Survey". In M. P. Kumar Reddy (Ed.), Mobile Information Systems.2022; 2022: 1–19. Hindawi Limited. Available: https: //doi. org/10.1155/2022/7519573.
[6] Kumar S, Kumar R. A Study on Different Aspects of Web Mining and Research Issues. In IOP Conference Series: Materials Science and Engineering.2021; 1022 (1): 012018. IOP Publishing. Available: https: //doi. org/10.1088/1757 - 899x/1022/1/012018.
[7] A. A. El - aziz, P. S. Pandian, S. N. Almuayqil, and A. S. Alruwaili, "A Framework for Clustering & Enhanced Approach for Frequent Patterns in Web Usage Mining A Framework for Clustering & Enhanced Approach for Frequent Patterns in Web Usage Mining Assistant Professor, Department of Information Systems, College of Computer, " no. June, 2020.
[8] S. Asadianfam, H. Kolivand, and S. Asadianfam, "A new approach for web usage mining using case based reasoning, " SN Appl. Sci., vol.2, no.7, pp.1–11, 2020, doi: 10.1007/s42452 - 020 - 3046 - z.
[9] P. M. Bharti and T. J. Raval, "Improving Web Page Access Prediction using Web Usage Mining and Web Content Mining, " Proceedings of the 3rd International Conference on Electronics and Communication and Aerospace Technology, ICECA 2019. pp.1268–1273, 2019, doi: 10.1109/ICECA.2019.8821950.
[10] R. Roy and G. Appa Rao, "Survey on pre - processing web log files in web usage mining, " Int. J. Adv. Sci. Technol., vol.29, no.3 Special Issue, pp.682–691, 2020.

Author Profile

Dr. S. Brindha received B. Sc degree in Science from Bharathiyar University. She done her M. Sc in Periyar University and she awarded M. Phil Computer Science from the Bharathiyar University. She received the Ph. D degree in Computer Science from the Bharathiar University. She has 8 years of teaching experience and 6 years of Technical Experience. At present she is working as Assistant Professor in Department of Computer Applications in SRM Faculty of Science and Humanities, SRM Institute of Science and Technology, Kattankulathur, Chengalpattu, Chennai, Tamilnadu, India. She published around 32 research papers in International Journals and Conferences. Published E - learning Concepts and Development tool book in the year of 2021. Published Exploring Taxonomy Based Methods for Detecting Patterns in Text Documents book in the year 2022. In the year 2023, published Data Science and Analytics. In the year 2024. She has published Information Security and Data Privacy. She have Published various book Chapters related to Climate Change and Human Health, Exploring Women Leadership: Achieving an Equal Future, Sustainable Solution for Green Environment. Received Women Researcher Award, Best Young Scientist Award and Best Faculty Award in Research. Member in International Research Awards on Science, Technology and Management. Lifetime Membership in Professional Nobel Sansnow's Professional Foundation, Approved by Ministry of Corporate Affairs, Government of India. Member in Computer Society of India and Association Computing Machinery (ACM). Her Research area includes Text Mining, Image Processing, Pattern Taxonomy Mining, Deep Learning, Artificial Intelligence and Machine Learning.