IJNRD.ORG ISSN : 2456-4184
International Research Journal IJNRD Research Through Innovation
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

Detection of Cyber bullying On Social Media Using Machine Learning

Disha Deotale
Department. of Comp & IT
G. H. Raisoni Institute of Engineering & Technology
(Savitribai Phule Pune University)
Pune, India

Priyanka Thorat
Department. of Information Technology
G. H. Raisoni Institute of Engineering & Technology
(Savitribai Phule Pune University)
Pune, India

Sara Kangane
Department. of Information Technology
G. H. Raisoni Institute of Engineering & Technology
(Savitribai Phule Pune University)
Pune, India

Pratiksha Yewale
Department. of Information Technology
G. H. Raisoni Institute of Engineering & Technology
(Savitribai Phule Pune University)
Pune, India

Sejal Indalkar
Department. of Information Technology
G. H. Raisoni Institute of Engineering & Technology
(Savitribai Phule Pune University)
Pune, India

Abstract:

Living in a digital world has both benefits and drawbacks. Web 4.0 also includes online abuse, a form of cybercrime. Technology is mostly used to harass others. We call this cyberbullying. This research paper looked at 30 different researchers' work on online bullying and how they found it. Cybercrime is any crime committed using the World Wide Web as an access point and an electrical gadget such as a computer or phone. Lack of datasets, information about predators, and victim privacy have all hampered cyberbullying research. A text mining method based on machine learning algorithms is proposed to detect bullying text before it occurs. Myspace.com and Perverted-Justice.com data were used to test the system. Compared to an initial study on the same dataset, these extract textual, behavioral, and demographic features. This goes beyond the previous study's focus on textual features. Words in the text could really result in cyberbullying if used. Cyberbullying may occur if these words are used in the text. If anyone bullies once, they may bully again. They give us personality traits. A dataset's demographic features include age, gender, and location. The device is tested by comparing the performance of both classifiers. The SVM classifier outperforms the Bernoulli NB with an overall accuracy of 87.14.

I. Introduction

Because of the tremendous increase in the availability of data services around the world, social media addiction in society has increased proportionally. India, like other countries, has seen a significant increase in cyberbullying. It is extremely difficult to protect society from the alarming rise in cyber-crime in this era of web 4.0, where people live on digital and online platforms. According to surveys, adolescents are the most common victims of cyberbullying. The following are examples of cyberbullying attacks carried out by attackers: (1) sending or posting hateful or abusive comments with the intent to harm an individual's character (2) Posting an offensive image or video. (3) Creating a false or inappropriate website. (4) Making online threats that cause someone to kill themselves or injure someone else. (5) Inciting religious, racial, ethnic, or political hatred online through the posting of hateful comments or videos.

II. LITERATURE SURVEY

“Rapid Cyber-bullying detection method using Compact BERT Models” Mitra Behzadi, Ian G. Harris, Ali Dera khshan” " Nowadays, many people use their social media platforms to spread hate online, which is why many researchers have focused on the problem of cyber-bullying detection over the last decade. Transfer learning is used to address this issue in this work. We employ a variety of 1compact BERT models and fine-tune those using data on hate speech. We incorporate the Focal Loss function to account for data class imbalances. We achieved state-of-the-art results of 0.91 precision, 0.92 recall, and 0.91 F1-score using this approach on the hate-speech dataset. Additionally, we demonstrate using our transfer learning pipeline that the more foldable BERT models are significantly faster at detecting cyberbullying and are suitable for real-time cyberbullying detection applications.

"A Review of Machine Learning Methods for Cyberbullying Detection in Social Media" Sanjay Kumar and Neha Singh Sharma. The modern period has witnessed a huge growth in internet usage, resulting in vast volumes of data being created. The positives and cons of the cyber world are numerous. One of the frightening circumstances in web 4.0 is cyberbullying, which is a sort of cybercrime. This article summarised the findings of different cyberbullying researchers and examined the different approaches used to identify cyberbullying as well as how to save society from online bad deeds like cyberbullying.

“A Fairness-Aware Fusion Framework for Multimodal Cyberbullying Detection “Jamal Alasadi, Ramanathan Arunachalam, Pradeep K. Atrey, Vivek K. Singh. Recent reports of bias in multimedia algorithms (e.g., lower face detection accuracy for women and people of color) have highlighted the critical need to develop approaches that work equally well for different demographic groups. As a result, we believe that ensuring fairness in multimodal cyberbullying detectors (e.g., equal performance regardless of victim gender) is an important research challenge. We suggest a fairness-aware alignment framework to ensure that when merging data from various modalities, both fairness and accuracy are important considerations. The inputs from different modalities are combined in this Bayesian fusion framework in a way that takes into account the different confidence levels associated with each feature as well as the interdependencies between features. This framework, in particular, assigns weights to different modalities based not only on accuracy but also on fairness. The results of applying the framework to a multimodal (visual + text) cyberbullying detection problem demonstrate the proposed framework's value in ensuring both accuracy and fairness.

“Text Imbalance Handling and Classification for Cross-platform Cybercrime Detection using Deep Learning” Munipalle Sai Nikhila, Aman Bhalla, Pradeep Singh. Cyberbullying has become a very common problem in the last few years. It's not just a problem for women. People of all ages are being bullied on any social media site. It is very important to build an artificial intelligence model that can inform 1if someone is being bullied in cross-platform posts. However, textual datasets that can be used to make models are very imbalanced. When textual data isn't balanced, we come up with two main ways to deal with it: synonym replacement and making new data using generative adversarial neural networks. We show how we used a convolutional neural network classifier to look at all of our methods. Work done by us shows that removing data imbalance with generative adversarial network techniques before classifying improves the model as a whole.

Cyberbullying is a persistent problem in Saudi schools, exacerbated by the advancement of digital technology and its pervasive presence in almost every societal aspect. With such technologies, it is unsurprising that harassment has spread to the virtual world of teenagers, where it is rampant. The intensity and outcome measures of this phenomenon have alarmed interested parties, but researchers who examined the causes and motivating factors behind cyberspace bullying participation are few and far between. 2The Theory of Planned Behavior, a well-known theory, was used to examine this issue (TPB). This study specifically looked at the effects of attitudes, normative beliefs, subjective norms, and perceived behavioral control/self-efficacy on cyberbullying intentions and expected societal outcomes. 1The study distributed 395 questionnaires to Saudi high school students from the ninth to the twelfth grades. The collected data were subjected to multiple linear regressions, with the results revealing that behavioral attitudes, social norms, perceived behavioral controls, social media use, a lack of parental controls, and a lack of regulations all had a direct effect on intentions to interact in cyberbullying. The findings also revealed that cyberbullying intentions had a direct impact on student academic performance. This study adds to our understanding of students' intentions toward cyberbullying and the relationship between the Theory of Planned (TPB) variables and the predictive utility model. Finally, the findings of this study can be used to develop prevention and intervention strategies, which have many implications for theory, practice, and policy.

Cybercrime refers to any crime committed using the internet as an access medium and using an electronic device such as a computer or a mobile phone. The main factors limiting previous research in cyberbullying detection have been a lack of datasets, predators' hidden identities, and victims' privacy. Taking these factors into account, an effective text mining approach based on machine learning algorithms is proposed for proactively detecting bullying text. The dataset gathered from myspace.com and Perverted-Justice.com was being used to assess the system's performance. When compared to a previous study on the same dataset that only considered textual features, three types of features are extracted from the dataset: textual, behavioral, and demographic features. Age, gender, and location are among the demographic features extracted from the dataset. The system is evaluated using various performance measures for both classifiers used, and the performance of the Support Vector Machine classifier is found to be better than the Bernoulli NB with an overall accuracy of 87.14.

“A Bag-of-Phonetic-Codes Model for Cyber-Bullying Detection in Twitter “ Ankita Shekhar, M. Venkatesan In today's digital world, social networking sites such as Twitter and Facebook are becoming increasingly important tools of communication. These things are now a normal part of life. With the help of their 1friends and family, people can openly discuss what they're thinking and how they're spending their time. However, there are some drawbacks to this freedom of expression. In some cases, people use social media to express their aggression, which in turn hurts the feelings of those they target. Sexual, racial, and physical disability-based cyberbullying are examples of this type of bullying. A thorough investigation is needed to deal with these kinds of situations. On a daily basis, Twitter as a micro-blogging site sees cyber abuse. Tweets, on the other hand, are unpolished texts, full of typos and censored language. A Bag-of-Phonetic-Codes model for detecting cyberbullying is proposed in this paper. Words can be corrected and censored by using their pronunciation as features. Reducing the vocabulary by correctly identifying duplicate words can help save feature space. The Bag-of-Words model, famous for extracting textual features, served as an inspiration for this project. The Soundex Algorithm has been used to generate phonetic codes. Experiments with both supervised and unsupervised machine learning approaches were conducted on a range of different datasets to better understand the approaches and challenges of cyberbullying detection.

“Cyber-Smart Children, Cyber-Safe Teenagers: Enhancing internet Safety for Children” Tracy WERU1, Joseph SEVILLA2, John OLUKURU3, Lorna MUTEGI4, Tabitha MBERI5 "Millennials have a lot more access to online resources than their parents did, and some teachers don't even know how to use them." Most of the time, children are better at technology than adults, so adults are afraid and don't enforce rules that are important for protecting their kids as they use resources online. Cybercrime is on the rise around the world, and Kenyan children are becoming victims because the country has a weak Children's Act 2001 that doesn't protect victims or punish offenders enough. A law that protects children from both physical and psychological abuse doesn't do enough to protect them from cybercrimes such as cyberbullying, pornography, lottery scams, and fake attacks. This paper shows how important it is to have a mobile game app for kids, teachers, and parents that teaches them about cyber safety.

“Identification of Potential Cyber Bullying Tweets using Hybrid Approach in Sentiment Analysis” Akankshi Mody, Shreni Shah, Reeya Pimple, Narendra Shekokar. Cyberbullying is on the rise because of the growth of the Internet. There is a pressing need to detect examples of cyber bullying due to the devastation it has caused. Use sentiment analysis to do just that. We use Natural Language Processing and Machine Learning techniques to detect cyberbullying in Tweets. If a tweet is flagged as a possible cyberbullying threat, it will be removed from the feed.

This paper looks at Ask.fm, a social networking site where users can make profiles and send each other questions. It looks at aggressive user behavior that could lead to cyberbullying. We think anonymity is a big reason why people act so aggressively on social networks. We look at how anonymous and not-anonymous people act on social networks. We took data from Ask.fm and looked at questions and answers that were posted by anonymous and non-anonymous users, as well as answers that were posted by non-anonymous users. Analysis of the data shows that anonymous users act more aggressively than non-anonymous users. Analysis also shows that people become more aggressive when they answer aggressive anonymous questions than when they answer aggressive anonymous questions that are not anonymous.

III. PROPOSED SYSTEM

Cyberbullying is a difficulty currently as a result of everybody using social media." Project aims to prevent cyberbullying. This proposal's goal is to boost the accuracy of cyberbullying detection by detective work grim tweets.

SYSTEM ARCHITECTURE

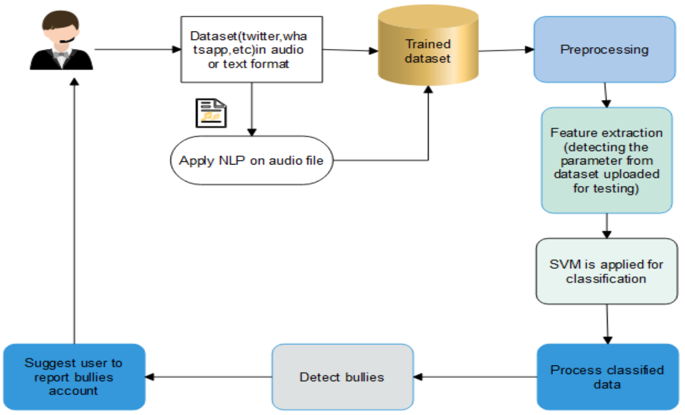
Dataset(twitter,wha tsapp,etc)in audio or text format
Trained dataset
Preprocessing
Apply NLP on audio file
Feature extraction (detecting the parameter from dataset uploaded for testing)
SVM is applied for classification
Suggest user to report bullies account
Detect bullies
Process classified data

Fig 1. System Architecture

IV. METHODOLOGY

SVM

Classifier is a famous Supervised Training algorithm used for Regression and Classification. Even so, it's often used through Machine Learning for Classification. The SVM algorithm's goal is to discover the best line or regression line that divides n-dimensional space into classes so that prospective data points can be easily classified. A hyper plane is the right decision dividing line.SVM selects the hyper plane's extreme points/vectors. Such severe situations are called support vectors, and the methodology is called SVM.

NLP

NLP algorithms are often based on machine learning. Instead of manually coding large sets of rules, NLP can use machine learning to learn them by analysing a set of examples (a book, for example) and making statistical inferences.

V. CONCLUSION

The above paper proposes a system for detecting Hindi and English tweets in Twitter. Because cyberbullying is highly dependent and contextual, sentiment and other contextual clues can help detect it. The system uses sarcastic tweets, not a data source of 9,104 cyberbullying tweets. The system uses LR. The approach has shown good results, with LR

]

classifier being more accurate than others. The extracted patterns do not cover all sarcastic detection patterns. The survey concluded that traditional machine learning algorithms cannot handle the massive amounts of data generated by Web 4.0, nor can cyber bullying content be accurately detected. Many researchers have recently become interested in Deep Learning, NLP, CNN, and stacked auto-encoder. Future research can use deep learning to detect cyberbullying in social media. Cyberbullying is a hotly debated topic. It is an emerging issue in Web 4.0. After reviewing 30 research papers, it was discovered that there is a lack of proper dataset, and that integrating social, contextual, and sentiment features can improve the monitoring of bullying content. Aside from text, images and video must be recommended for future career

VI. REFFERENCES

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

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

6) 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) 7 V. Nahar, X. Li, C. Pang, and Y. Zhang. ”Cyberbullying detect ion based on text-st ream classification.” In The 11th Aust ralasian Data Mining Conference (AusDM 2013), 2013.

8) 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.

International Research Journal
IJNRD
Research Through Innovation