International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

DATA SECURITY APPROACH ON CYBER CRIME WITH WEB VULNERACALBILITY

A PAVANI[1], S.K.ALISHA[2]

[1]MCA Student, B V Raju College, Kovvada, Andhra Pradesh, India.
[2]Associate Professor, B V Raju College, Kovvada, Andhra Pradesh, India.

ABSTRACT:

A crime is a deliberate act that can cause physical or psychological harm, as well as property damage or loss, and can lead to punishment by a state or other authority according to the severity of the crime. The number and forms of criminal activities are increasing at an alarming rate, forcing agencies to develop efficient methods to take preventive measures. In the current scenario of rapidly increasing crime, traditional crime-solving techniques are unable to deliver results, being slow paced and less efficient. Thus, if we can come up with ways to predict crime, in detail, before it occurs, or come up with a "machine" that can assist police officers, it would lift the burden of police and help in preventing crimes. To achieve this, we suggest including machine learning (ML) and computer vision algorithms and techniques. In this paper, we describe the results of certain cases where such approaches were used, and which motivated us to pursue further research in this field. The main reason for the change in crime detection and prevention lies in the before and after statistical observations of the authorities using such techniques. The sole purpose of this study is to determine how a combination of ML and computer vision can be used by law agencies or authorities to detect, prevent, and solve crimes at a much more accurate and faster rate. In summary, ML and computer vision techniques can bring about an evolution in law agencies. Social networking and online chatting application provide a platform for any user to share knowledge and talent but few users take this platform to threaten users with cyberbullying attacks which cause issues in using these platforms.

Keywords: ML, Crime, cyberbullying, predict crime.

1. INTRODUCTION

Web 2.0 is extending and evolving in terms of the volume, velocity and variety of information accessible online across various social media portals which affirm that the social media (SM) has global reach and has become widespread [1]. The global and pervasive reach of

social multimedia has in return given some unpremeditated consequences where people have discovered illegal & unethical ways to use the socially connected virtual communities. There are various benefits of SM but few people use it in wrong way. One of its most severe upshots is known as cyberbullying where individuals find new means to bully one another over the Internet. The term „Cyberbullying" was devised by anti-bullying activist Bill Belsey in the year 2003 [2]. Tokunaga defined cyberbullying as "any behavior performed through electronic or digital media by individuals or groups that repeatedly communicates hostile or aggressive messages intended to inflict harm or discomfort on others" [3]. Different characteristics are highlighted by this definition like, the technology part, the antagonistic behaviour of the act, important reason for causing suffering, considered very crucial to the definition and repetitiveness by many of the scholars [1]. Cyberbullying over social networks has already been claimed as a major risk or threat. Cyberbullying can be possible through any of the media like mobile phones or using internet. Cyberbullying may be done through emails, instant messages, chat room, blogs, images, video clip, text messages etc. [4, 5]. It has grown as a social menace that puts a negative effect on the minds of both the victim and bully. It is more persistent way of bullying a person before an entire online network, that is, the social networking sites, which can ultimately result in emotional and psychological breakdown of the victim with developed feelings of depression, stress, lack of self-confidence, anger, sadness, loneliness, health degradation, and suicides etc.

Literature survey:

Today, because of enormous development of web2.0, online presence of individuals is normal and permanent. Additionally, the risk of cyberbullying and the pessimism brought about by cyberbullying is expanding. Hence a lot of research is currently being done around there, particularly for Cyberbullying Detection. Recent literature accounts the use of supervised machine learning and deep learning techniques for classifying hate speech, aggression, comment toxicity and bullying content on social forums [14, 15]. A portion of the work that offers

sight to this issue is done by scientists in [16-18]. Theoretical aspects of cyberbullying and how it is prevailing among youths and youngsters have been examined in [19]. The characteristics profile of wrongdoers and victims and conceivable strategies for its preventions are introduced in [18]. Till now the majority of the work is devoted to text analysis for the most part. The work done in [9, 19, 20-24] basically is research on Cyber-Aggression that has utilized text investigation approach on the comments. Work done in [22, 25-27] uses text-based analysis for identification of cyberbullying utilizing the dataset from formspring.me and Myspace. Dinakar et al. in [19] construct a Bully Space, a common-sense knowledge base that encodes particular knowledge about bullying situations and analyse the messages on Form spring (a social networking website) using Analogy Space common sense reasoning technique. Hinduja et al. [17] explored the relationship between cyberbullying and suicide among adolescents. The characteristic profile of wrongdoers and victims and conceivable strategies for its preventions are introduced in [18]. Till now the majority of the work is devoted to text analysis for the most part. The work done in [19-24] basically is research on Cyber-Aggression that has utilized text investigation approach on the comments. Work done in [22, 25-27] uses text-based analysis for identification of cyberbullying utilizing the dataset from Formspring.me and Myspace.

PROPOSED SYSTEM

The proposed deep classification model reinforces the strengths of deep learning nets in combination to machine learning to deal with different modalities of data in online social media content. The proposed CNN-BoVW-SVMmodel consists of four modules, namely, text analytics module, image analytics module, discretion module and decision module. The basic architecture of the work done has been clarified beneath (Fig 3.1). The steps included can be comprehended as 1. Analysing the sort of information. 2. Passing it to the separate module for processing. 3. Decision module is utilized to analyse the outcome. Analysing the sort of information includes checking whether the input is just text or it is a picture or it is a picture with text embedded on it. This is vital in

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

light of the fact that once we have investigated this then we can perform further handling in the respective modules. Text only: On the off chance that the input is as just text, at that point we will perform pre processing of text, extract the features, create the feature vector and after that utilization CNN is used for performing the task of classification. Image only: If the input is image only, at that point we will perform the pre-processing of image like converting it to Gray scale or resizing, then extricate features utilizing BoVW approach, produce histogram and after that uses SVM for doing the classification of the image. Image with Text: If the input is the image with text embedded on it on it, at that point an additional step will be included to separate that text from the picture, which we are doing utilizing Google Photos as a tool. When we have the text separated from the image, we can utilize the means utilized for performing text analysis and for the picture we will utilize the image handling steps. The result of those two modules will be encouraged as contributions to a Boolean framework that will at that point shows the outcome whether it is a bully or not. When one of the inputs to Boolean framework isn't accessible like we have only text or in the event that we have picture just, at that point that input to Boolean framework will be unfilled or false since we are utilizing an OR operation in the Boolean framework, if the text or image is a sort of bully, it will get identified.

2. AN OVERVIEW OF PROPOSED SYSTEM

Classifier:

Naive Bayes:- This classifier belongs to the probabilistic group of classifiers in the domain of machine learning. The base of this classifier is the Bayes Theorem where the features are considered to be independent of each other. It is a very popular when it comes to classification. It is a simple model where the test (unknown) instances are assigned class tags based on the trained model.

K-NN :- K-nearest neighbour model can be used as classification model or regression model. For an unclassified instance as the input, we consider the k classified instances in a constraint region and accordingly the unclassified instance is given a class whose instances are most in that region. In case K=1, the

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

unclassified instance is given the class whose neighbour is nearest to it, there is no need for count as the value of k is 1.

SVM :- A Support Vector Machine (SVM) works by finding a hyper-plane that can efficiently divide the set of objects in different classes.SVM takes a labelled training data, and outputs an optimal hyper-plane which can then be used to categorize new examples. A decision plane separates set of objects having memberships of different classes. For a 2d space, this hyper-plane or decision boundary is a straight line. In this image analytic module, SVM analyses data and recognizes image patterns. A set of training examples is provided to the algorithm and it generates a boundary in order to differentiate between the classes learning from training examples.

SMO: - Sequential minimal optimization helped the support vector machine (SVM) with the problem of quadratic programming. It was developed at the Microsoft Research in 1988 by John Platt. SMO is used in the training phase of the SVM so as to get rid of the problem. It was quiet an important development as in early days if was very expensive to get rid of the quadratic programming problem of SVM using 3-party software.

Random Forest: - It is also known as - Random decision forests, it is an ensemble learning technique used for both regression and classification. It works by generating large number of decision trees in the training phase and in the test, phase gives the result according to whether it is for classification or regression. It is better than decision tress as it removes its limitation of getting too precise depending on the training dataset. Its first creation was done by Tin Ham Ho in the year 1995.

Fig.1. Home page.

Fig.2. Registration form.

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

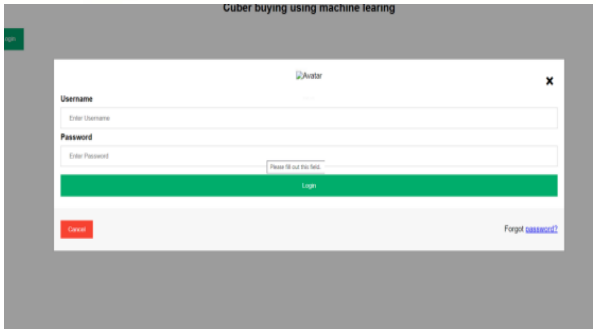

Fig.3. Admin page.

Fig.4. User details.

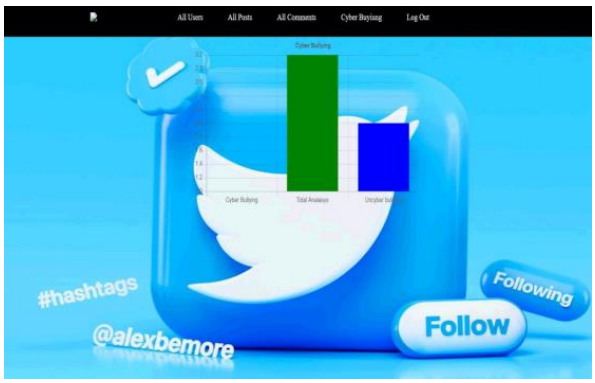

Fig.5. User details.

3. CONCLUSION

Social media and the internet have opened up new forms of both empowerment and oppression. Meaningful engagement has transformed into a detrimental avenue where individuals are often vulnerable targets to online ridiculing. Predictive models to detect this cyberbullying in online content is imperative and this research proffered a prototype model for the same. The uniqueness of the proposed hybrid deep learning model, CNNBoVW-SVM is that it deals with different modalities of content, namely, textual, visual (image) and info-graphic (text with image). The results have been evaluated and compared with various baselines and it is observed the proposed model gives superlative performance accuracy. The limitations of the model arise from the characteristics of realtime social data which are inherently „high-dimensional", „imbalanced or skewed", „heterogeneous", and „cross-lingual". The growing use of micro-text (wordplay, creative spellings, slangs) and emblematic markers (punctuations and emoticons) further increase the complexity of real-time cyberbullying detection.

FUTURE SCOPE

One way to alleviate the diversity of bullying is to find other ways in labeling. In this thesis, we just label bullied/non-bullied posts by going through the comments, pictures or memes which would be affected by commenters" preferences. To give a simple example, in sports, fans tend to verbally attack their team"s rivals whenever they get a chance.

In our situation, a photo in which a football player is visiting a children‟s hospital might receive some offensive comments just because there are supporters of his rival teams among the viewers. Yet the image probably looks innocent to other viewers if they are not aware of who the person is. If we can increase the number of labelers for each image and ask them to label the instances only after viewing the photos themselves based on their own subjective opinions, the diversity of bullying could be mitigated. Besides that, we should design or look for features that can more efficiently recognize the factors that cause viewers to bully. Emotional information from the posts would be a possible choice to implement since the bully is a product of extreme emotions. Also, strong object detection techniques could be considered.

REFERENCES

[1]. Aboujaoude E, Savage MW, Starcevic V, Salame WO. Cyberbullying: review of an old problem gone viral. J Adolesc Health. 2015;57(1):10–18. doi: 10.1016/j.jadohealth.2015.04.011.

[2]. Campbell MA (2005) Cyber bullying: An old problem in a new guise? Journal of Psychologists and Counsellors in Schools 15(1):68-76.

[3]. Tokunaga Following you home from school: a critical review and synthesis of research on cyberbullying victimization. Comput Hum Behav. 2010; 26:277–287. doi: 10.1016/j.chb.2009.11.014.

[4]. Centers for Disease Control and Prevention. Youth violence: technology and youth protecting your child from electronic aggression; 2014. http://www.cdc.gov/violenceprevention/pdf/ea-tipsheet-a.pdf. Accessed 11 September 2017.

[5]. Smith PK, Mahdavi J, Carvalho M, Fisher S, Russell S, Tippett N. Cyberbullying: its nature and impact in secondary school pupils. J Child Psychol Psychiatry. 2008;49(4):376–385. doi: 10.1111/j.1469-7610.2007.01846.

[6]. Hinduja, S. & Patchin, J. W. (2014). Cyberbullying Identification, Prevention, and Response. Cyberbullying Research Center (www.cyberbullying.us)

[7]. https://machinelearningmastery.com/best-practices-document-classification deeplearning/

[8]. Dadvar, Maral, and Kai Eckert. "Cyberbullying Detection in Social Networks Using Deep Learning Based Models; A Reproducibility Study." arXiv preprint arXiv:1812.08046(2018).

[9]. K. Dinakar, R. Reichart, and H. Lieberman. Modeling the detection of textual

cyberbullying. In The Social Mobile Web, 2011(21, 30).

[10]. Hosseinmardi, Homa, et al. "Prediction of cyberbullying incidents in a mediabased social network." 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM). IEEE, 2016).