Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 16307 - 16315
Received 05 March 2021; Accepted 01 April 2021.

Detecting Twitter Cyberbullying Using Machine Learning

Dr.A.K.Jaithunbi[1], Gollapudi Lavanya[2], Dondapati Vindhya Smitha[3], Bandi Yoshna[4]

[1]Assistant Professor,Department of Computer science and Engineering, R.M.D.Engineering College, Thiruvallur,India
[2]Under-Graduate Student, Department of Computer Science and Engineering, RMD Engineering College, Thiruvallur, India.
[3]Under-Graduate Student, Department of Computer Science and Engineering, RMD Engineering College, Thiruvallur, India.
[4]Under-Graduate Student, Department of Computer Science and Engineering, RMD Engineering College, Thiruvallur, India.
akj.cse@rmd.ac.in[1], lavanyagollapudi5166@gmail.com[2], vsmitha27@gmail.com[3]

Abstract—

Online media is a stage where numerous youthful individuals are getting tormented. As person to person communication destinations are expanding, cyberbullying is expanding step by step. To recognize word likenesses in the tweets made by menaces and utilize AI and can build up a ML model naturally recognize online media tormenting activities. In any case, numerous online media tommenting identification methods have been actualized, however numerous of them were printed based.

The objective of this paper is to show the execution of programming that will distinguish tormented tweets, posts, and so on An AI model is proposed to distinguish and forestall tormenting on Twitter. Two classifiers for example SVM and RF are utilized for preparing and testing the online media tormenting content. Both SVM (Support Vector Machine) and RF had the option to recognize the genuine positives with 71.25% and 52.70% precision individually. Yet, SVM beats RF of comparable work on the equivalent dataset.

Keywords: Machine learning techniques, naive bayes, random forest, logistic regression, Support Vector Machine, Sentiment Analysis.

I. INTRODUCTION

Nowadays technology has become a very important part of our lives and most people can't live without it. The Internet provides a platform to share their ideas. Many people are spending a large amount of time on social media. Communicating with people is no exception, as technology has changed the way people interact with a broader manner and has given a new dimension to communication. Many people are illegally using these communities. Many youngsters are getting bullied these days. Bullies use various services like Twitter, Facebook and Email to bully people. Studies show that about 37% of children in India are involved in cyber bullying and nearly 14% of bullying occurs regularly. Cyber

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 16307 - 16315
Received 05 March 2021; Accepted 01 April 2021.

bullying affects the victim both ways emotionally and psychologically. Social media also allows bullies to harness the anonymity which satisfies their unkind deeds. Things also get more serious when bullying occurs more repeatedly over time. So, preventing it from happening will help the victim.

[2]. Cyber bullying is an act of threatening, harassing or bullying someone through modern ways of communicating with each other and with anybody/everybody in the world via social media apps/sites. Cyber bullying is not just limited to creating a fake identity and publishing/posting some embarrassing photo or video, unpleasant rumors about someone but also giving them threats. The impacts of cyber bullying on social media are horrifying, sometimes leading to the death of some unfortunate victims. The behavior of the victims also changes due to this, which affects their Emotions, self-confidence and a sense of fear is also seen in such people.

Thus, a complete solution is required for this problem. Cyber bullying needs to stop. The problem can be tackled by detecting and preventing it by using a machine learning approach, this needs to be done using a different perspective.

The main purpose of our paper is to develop an ML model so it can detect and prevent social media bullying, so nobody will have to suffer from it. The proposed technique is implemented on the social media bullying dataset which was collected from various sources like Kaggle, GitHub, etc.

The performance of both NB and SVM is compared to TFIDF. Twitter API is used to fetch a particular location's tweets to detect whether they are Bullying or not. Furthermore, the probability of each tweet is calculated to predict the result and the result of each tweet is stored into the database with bullies' username.

II. LITERATURE SURVEY

1.Under the guidance of software engineering theory, this paper designs a Graduation Thesis Quality Management System based on cloud platform by Web programming technology. The system organically integrates Internet technology into graduation thesis management, and effectively regulate all aspects of graduation thesis management. The process management and quality control of graduation thesis are realized, the management efficiency is improved, and the informationization, networking and standardization of graduation thesis management all can be realized.

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 16307 - 16315
Received 05 March 2021; Accepted 01 April 2021.

2.With the popularity of online education and e-learning, it is necessary for universities to realize the online management of graduation thesis, which is one of the most important teaching processes. In this paper, the technology solution including JSP + Tomcat + SQL Server 2005 is used to implement a web-based system for graduation thesis management, which can standardize the whole management process from thesis guide to oral defense. Based on requirements and feasibility analysis, technical solutions, system structure and core functions of this online system are introduced. Then, focusing on the system implementation of some core functionality, the key technologies are analyzed in detail.

3. After analyzing most of the current Graduation Thesis Management Systems, we proposed an intelligent management system which can meet the requirements of the process of most Graduation Thesis Management Systems. In this paper we mainly introduce the design of the functional components and implementation of the system. We developed the system using ASP.NET and SQL Server database and designed user interfaces for the system manager, principal, student, advisor and supervisor. The system supports batch lead-in of teachers and students' information, two-way choice of thesis topics, managements to thesis grades, statistics and export of thesis related information as well as automated analysis to thesis topic's repetition frequency

4.The existing ERP system of the existing enterprise management system so that the management level to stay in the local network level, cannot form information exchange with the outside world, hinder the interaction between enterprises inside and outside the network information, the author proposed a based on the Internet plus and cloud platform progress monitoring system, taking full account of the internal and external network data transmission based on the security, filtering the enterprise sensitive information, in order to achieve data entry, progress query and early warning three functions.

5.Ubiquitous healthcare services are becoming more and more popular, especially under the urgent demand of the global aging issue. Cloud computing owns the pervasive and on demand service-oriented natures, which can fit the characteristics of healthcare services very well. However, the abilities in dealing with multimodal, heterogeneous, and non stationary physiological signals to provide persistent personalized services, meanwhile keeping high concurrent online analysis for public, are challenges to the general cloud. In this paper, we proposed a private cloud platform architecture which includes six layers according to the specific requirements. This platform utilizes message queue as a cloud engine, and each layer thereby achieves relative independence by this loosely coupled means of communications with publish/subscribe mechanism. Furthermore, a plug-in algorithm framework is also presented, and massive semi-structure or unstructured medical data are accessed adaptively by this cloud architecture.

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 16307 - 16315
Received 05 March 2021; Accepted 01 April 2021.

As the testing results showing, this proposed cloud platform, with robust, stable, and efficient features, can satisfy high concurrent requests from ubiquitous healthcare services.

III. EXISTING SYSTEM

Twitter is listed as one of the top five social media platforms where the maximum percentage of users experience cyberbullying (turbofuture.com, 2019). It enables a user to send a message of 280-characters, with more than 330 million active users at present (Statista, 2018). Studies on cyberbullying and Twitter often reported extensive cases of the phenomenon, with the potential for serious, deleterious consequences for its victims (Chatzakou et al., 2017a; Balakrishnan et al., 2019; Sterner, 2017). Several measures have been taken by Twitter to mitigate cyberbullying, such as filtering unwanted messages from users without a profile picture, and enabling a time-out feature that bans users using abusive language, among others. Despite these positive attempts, the platform is not completely immune from cyberbullying (Bernazzani, 2017; Twitter, 2019).

Sentiments are the thoughts or opinions provoked due to the feelings attached with something, often categorized as positive, neutral or negative (Zhao et al., 2016). The process in which the unstructured data are computationally processed is referred to as sentiment analysis, and can be categorized into machine learning approach, lexicon-based approach and hybrid approach (Medhat et al., 2014). The machine learning approach employs algorithms such as Naïve Bayes, SVM, Decision Tree, etc., whereas the lexicon based approach is dependent on sentiment lexicons (i.e., dictionary of opinion words and phrases with the assigned polarities and intensities) for gauging the sentiment of a text. In the context of cyberbullying, sentiment has been used to distinguish between bullies, victims and non-bullies (Dani et al., 2017; Nahar et al., 2012; Xu et al., 2012). For instance, (Xu et al. 2012) identified cyberbullying victims using their sentiment scores as victims usually experience negative emotions such as depression, anxiety and loneliness.

DISADVANTAGES

1. Accuracy will be low. The time complexity is very high because we are working on text data.

2. Covert the text data into numeric form is very big task. We have to text preprocessing like removing the stop words, punctuation marks...etc.

3. Time consuming and prediction is not perfect

IV. METHODOLOGY

The proposed application should be able to detect the bullyed tweets. The feature representation is done by logistic regression, to convert To implement the model the steps to be followed are

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 16307 - 16315
Received 05 March 2021; Accepted 01 April 2021.

- Dataset preparation and preprocessing
- Featuraization
- Data splitting
- Modeling Evaluation
- Hyper parameter Tuning
- Model Testing

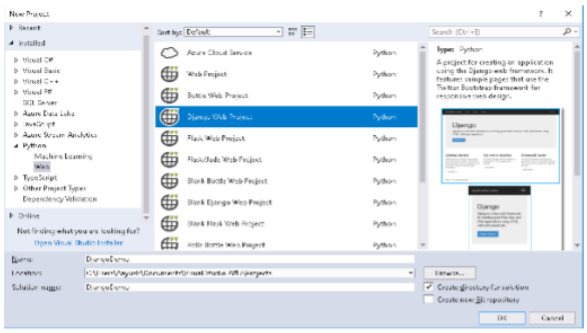
V. PROPOSED SYSTEM

In this paper, a solution is proposed to detect twitter cyberbullying. The main difference with previous research is that we not only developed a machine learning model to detect cyberbullying content but also implemented it on particular locations real-time tweets using Twitter API. The entire approach to detect and prevent Twitter cyberbullying is divided into 2 major stages: developing the model and experimental setup.

Stepwise Procedure of SVM and Naïve Bayes utilized in detecting the cyberbullying Steps:

1. For a particular location, a limited number of tweets will be fetched through Twitter's tweet API
2. The Data Preprocessing, Data Ext reaction will be performed on the fetched Tweets
3. Preprocessed tweets will be passed to SVM and Naïve Bayes model (see Developing the Model section) to calculate the probabilities of fetched tweets to check whether a fetched tweet is bullying or not.
4. If the probability of fetched tweet lies in the range of 0 to 0.5, then the tweet will not be considered as a bullied tweet. If the probability of the fetched tweet is above 0.5, it will be added to the database and then further 10 tweets from that users' timeline will be fetched, becauseit cannot directly say the person is bullying someone or not because it is might possible he's having a conversation with his friend hence to make sure whether he was bullying someone or not we will fetch last 10 tweets from his timeline and preprocessing will be performed over the tweets.
5. Again, the list of user's time line tweets will be passed to the SVM and Naive Bayes model to predict the results of the tweets.
6. And again, the average probability of that user's tweets will be calculated and if it lies above 0.5 then it will be considered as a bullied tweet and it will be recorded in our database. If the average probability is less than 0.5 then the record will be removed from the database.

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 16307 - 16315
Received 05 March 2021; Accepted 01 April 2021.



VI. ADVANTAGES

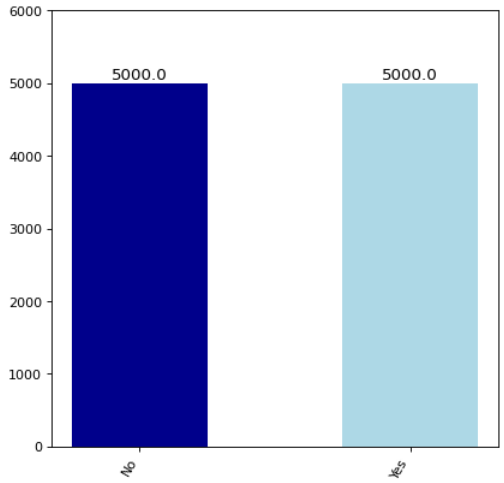
\

1. Cyberbullying research has often focused on detecting cyberbullying ‘attacks’ and hence overlook other or more implicit forms of cyberbullying and posts written by victims and bystanders. However, these posts could just as well indicate that cyberbullying is going on
2. Speed and very low complexity, which makes it very well suited to operate on real scenarios.
3. Using random forest or Liner SVM we can predict the cyberbullying very accurately.

VII. FEATURE EXTRACTION

In this step, the proposed model has transformed the data in a suitable form which is passed to the machine learning algorithms. The TFDIF vectorizer is used to extract the features of the given data. Features of the data are extracted and put them in a list of features. Also, the polarity (i.e. the text is Bullying or Non-Bullying) of each text is extracted and stored in the list of features.

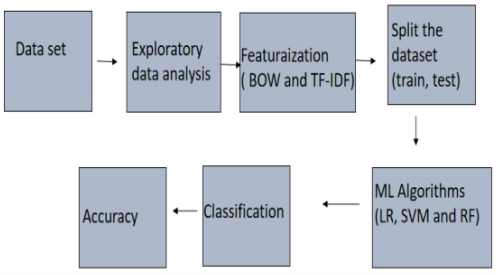
Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 16307 - 16315
Received 05 March 2021; Accepted 01 April 2021.



VIII. ALGORITHM SELECTION

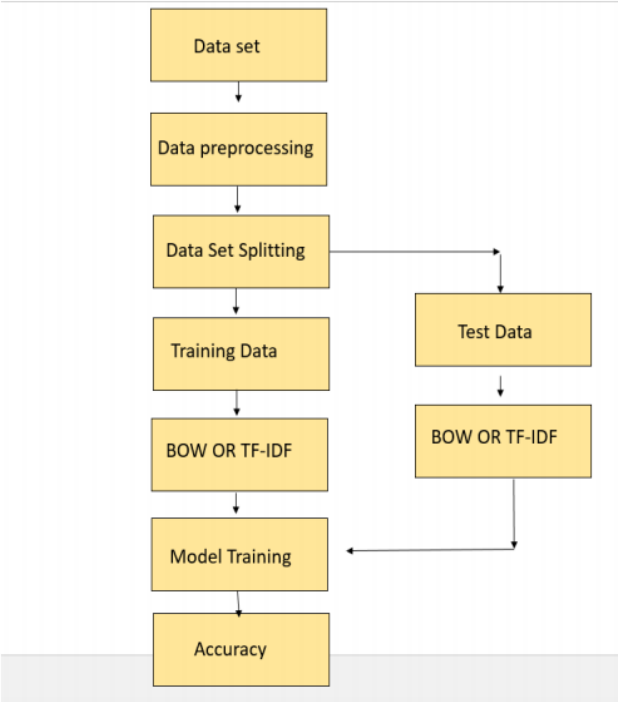
To detect social media bullying automatically, supervised Binary classification machine learning algorithms like SVM with linear kernel and Naive Bayes is used. The reason behind this is both SVM and Naive Bayes calculate the probabilities for each class (i.e. probabilities of Bullying and Non-Bullying tweets). Both SVM and NB algorithms are used for the classification of the two-cluster. Both the machine learning models were evaluated on the same dataset. But SVM outperformed Naive Bayes of similar work on the same dataset. Classification report is also evaluated. The accuracy, recall, f-score, and precision are also calculated. Precision = TP / (TP+FP) Recall =TP/ (TP+FN) F-Score = 2*(Precision*Recall) / (Precision + Recall) Where TP = True positive numbers

IX. BACK END MODULE DIAGRAM



x. State Diagram

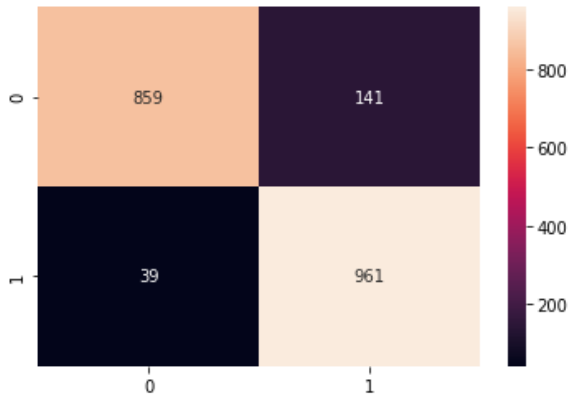
Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 16307 - 16315
Received 05 March 2021; Accepted 01 April 2021.



XII. CONCLUSION

The goal of this project is to the automatic detection of cyberbullying-related posts on social media. Given the information overload on the web, manual monitoring for cyberbullying has become unfeasible. Automatic detection of signals of cyberbullying would enhance moderation
and allow to respond quickly when necessary. However, these posts could just as well indicate that cyberbullying is going on. The main aim of this project is that it presents a system to automatically detect signals of cyberbullying on social media, including different types of cyberbullying, covering posts from bullies, victims and bystanders.

XII. RESULT



Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 16307 - 16315
Received 05 March 2021; Accepted 01 April 2021.

XIII. REFERENCES

[1] John Hani Mounir, Mohamed Nashaat, Mostafa Ahmed, Zeyad Emad, Eslam Amer, Ammar Mohammed, “Social Media Cyberbullying Detection using Machine Learning”, (IJACSA) International Journal of Advanced Computer Science and Applications Vol. 10, pages 703-707, 2019.

[2] Kelly Reynolds, April Kontostathis, Lynne Edwards, "Using Machine Learning to Detect Cyberbullying”, 2011 10th International Conference on Machine Learning and Applications volume 2, pages 241–244. IEEE, 2011

[3] Amanpreet Singh, Maninder Kaur, "Content-based Cybercrime Detection: A Concise Review”, International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue-8, pages 1193-1207, 2019

[4] Abdhullah-Al-Mamun, Shahin Akhter, "Social media bullying detection using machine learning on Bangla text”, 10th International Conference on Electrical and Computer Engineering, pages 385-388, IEEE Xplore, 2018

[5]NektariaPotha and ManolisMaragoudakis. “Cyberbullying detectionusing time series modeling”, In 2014IEEE International Conference on, pages 373–382. IEEE, 2014.

[6] Ying Chen, Yilu Zhou, Sencun Zhu, and Heng Xu. “Detecting offensive language in social media to protect adolescent online safety”. In Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on and 2012 International Conference on Social Computing (SocialCom), pages 71– 80. IEEE, 2012

[7] Vikas S Chavan, SS Shylaja. “Machine learning approach for detection of cyber-aggressive comments by peers on social media network". In Advances in computing, communications, and informatics (ICACCI), 2015 International Conference on, pages 2354–2358. IEEE,2015

[8] Walisa Romsaiyud, Kodchakorn na Nakornphanom, Pimpaka Prasert- silp, Piyaporn Nurarak, and Pirom Konglerd, “Automated cyberbullying detection using clustering appearance patterns”, In Knowledge and Smart Technology (KST), 2017 9th International Conference on, pages 242–247. IEEE, 2017.

[9] https://muthu.co/understanding-the-classification-report-in-sklearn/

[10] https://developer.twitter.com/en/apps

[11] https://text-processing.com/demo/tokenize/

[12]https://towardsdatascience.com/support-vector-machine-introduction-to-machine-learning-algorithms-934a444fca47

[13] https://towardsdatascience.com/naive-bayes-classifier-81d512f50a7c