Detection of Negative tweets with offensive words for Cyberbullying

[1]Prerana Gandhi, [2] Mr. Chandresh D Parekh

[1]Research Student, [2]Assistant Professor(RSU- Ahmedabad)

[1]M.tech Cyber Security,

[1]Raksha Shakti University, Ahmedabad, India

Abstract : In this work, we tackle one of the problems identified above, that of cyberbullying, which is continuously increasing in the social Web and is becoming a major threat to teenagers and adolescents. In fact, 55% of teens using social media have witnessed outright bullying via that medium. in the digital age. there has been increasing in cyberbullying against special community including the women and minorities [1]. Our project focuses on these two issues to detect bullying through test and sentiment analysis in tweets of Twitter. We use a dictionary of weighted offensive words along with the presence of pronouns or collective nouns to calculate the Offensiveness percentage in each tweet. We also make use of a Sentiment Analysis to detect sentiment of the tweets.

Keywords: cyberbullying, tweets, pre-processing, big-data, twitter, sentiment analysis

1.INTRODUCTION

People share their lives over social media and interconnect on daily basis this is the extent of reach in technology today. This has resulted in a vast quantity of data to be generated. In this particular lots of data, social media plays an important role especially sites like Facebook, Twitter, blogs and so on. With such large amounts of data, scalability, and processing such data with complex algorithms are challenging task. The field of data mining presents a task of lots of unique information. It also presents the need for inquisitive ways of gathering and processing of the data. In a way, it provides new insights ways into the data from the social network sites.The information gathered from this sites can be of various ways like personal information, their locations or preferences and analysis of text based on sentiment or words. Analysis of sentiment on the various subject shows the judgment of people, this is specially done in microblogging as in twitter. Using such sentiment and analysis of tweets with natural language text processing we can figure out if the tweet inoffensive or not. This is particularly helpful in finding out cyberbullying.

This idea of analysis of sentiments and tweets can be formed as a forensic subject now taken into forensic linguist as a subject.

2. RESEARCH SCENARIO

Twitter has vast potential for data mining as its users produce Big Data that can be processed for various purposes. In addition, there are requirements for architecture development that can scale to continuous new-streamed tweets and also able to integrate with advanced machine learning algorithms. Knowing what users think or how they feel about products is a value proposition for companies. Sentiment analysis is part of data mining, which monitors public perceptions about various topics. It can track what people think about business products and their quality, brands, pricing system or worldwide trends. Moreover, it can analyze business opportunities and thus become an effective factor for multinationals to innovate their services. Twitter as microblogging platform backed with its active users create opportunities for data mining and more particular sentiment analysis based on tweets. Twitter users often express their opinions and views about various topics within their posted tweets. Therefore, by using text-based processing data mining technique. It can serve to those multinationals as feedback or for brand management or product management. On the other hand, since Twitter generates the very large amount of data every day, sentiment analysis can be helpful with marketing related campaigns to research public opinions. for example blockbuster movie and analyze sentiment about users satisfaction, whether they felt positive or negative about the movie. According to consumers are willing to pay from 20% to 99% more for a movie rated with 5/5 stars [1]. This research discusses that positive/negative comments or reviews on product/services are great influencers and will indicate success among users.

3.PROBLEM UNDERSTANDING

Synopsis:

the world has seen many revolutions that were made possible by Social Media. It is an extremely influential innovation of our time and is a great way to expand the boundaries of one's experiences and become socially active. However, social media is a double-sided weapon.it can be either good for business and bad for cyberbullying

As the size of Twitter© data is increasing, so are undesirable behaviors of its users. A lot of antisocial behavior is observed on social media, including cyber-stalking, cyber-bullying, and cyber-harassment .one of the major undesirable behavior is cyberbullying, which could lead to catastrophic consequences.Moreover, this is not limited to children and young adults; anybody can be a victim. Hence, it is critical to efficiently detect cyberbullying behavior by analyzing tweets, in real time if possible. Prevalent approaches to identifying cyberbullying are mainly stand-alone, and thus, are time-consuming.

I.UNDERSTANDING CONCEPT OF CYBERBULLYING:

"Cyberbullying is formally defined as willful and repeated harm inflicted through the use of computers, cell phones, and other electronic devices"[4].In short, bullies typically exploit the use of electronic communication for harassing people.

This harassment may be motivated by anger, frustration, revenge, or from a basic desire to control others and feel more powerful (Why do kids, n.d.). Sometimes kids cyberbully others to cope with their own low self-esteem and/or to fit in with their peers (Why Do People, n.d.). Examples of cyberbullying can include rumors sent by e-mail or posted on social media; embarrassing pictures or videos; and intimidating, insulting, and/or harassing messages posted on social networks. Once such derogatory messages, pictures, or videos are posted, it is very difficult to take these posts off the social media sites.

It can happen 24 hours a day and 7 days a week, and it can even reach its victim when they are alone, outside in the schoolyard, or in the sports field [Patchin J. W., 2014].

Cyberbullying empowers a bully to humiliate and hurt the victim in online communities without ever getting recognized. Furthermore, the fear of getting socially cutoff stops the victims and people knowing about the incidents from reporting the incidents. Thus it becomes a tough challenge to control. Cyberbullying can lead to ruinous consequences like depression, sometimes death of the victim.

Studies performed by The Journal of Psychosocial Research on Cyberspace show that "critical impacts occurred in almost all of the respondents' cases in the form of lower self-esteem, loneliness and disillusionment and distrust of people: The more extreme impacts were self-harm and increased aggression towards friends and family" [Šleglova, 2011].

It also states that some victims develop "coping strategies". It is known that victims try to deal with cyberbullying themselves leading to a stressful situation. Parents find it tough to know that their child is cyberbullied is also a problem. The support system is able to help the victim after they identify the cyberbullying or signs of onset, as a victim do not approach directly. Hence a need for programs which detect cyberbullying and inform the representative authorities such as parents, support system etc.

II.Countermeasures set by twitter:

Social networks provide some degree of support for the safe web experience. Tools that help to protect one's privacy are as follows: Twitter provides users with the following tools ("Learn How", 2017).
1. Allowing users to block, mute, or unfollow unwanted followers.
2. Filters on notifications that allows users to filter out any unwanted replies or mentions from the accounts that the user do not follow.
3. Reporting the undesirable behavior to Twitter.
4. Warning the user about sensitive content before showing it. It works only for photos and videos.
5. Tagging privacy for photos allows the user to decide who can and cannot tag him/her in photos.

III.Challenges and Changes to be made

Although tools are provided by contemporary social networking sites and laws are in place to fight cyberbullying in some countries. the majority of cyberbullying instances go unreported (Peterson, 2013). At the same time, there is no system in place for automatic detection of such behavior. Cyberbullying is one of the widely recognized problems which has a lasting impact on its victims. While healthy social behavior is the solution to this problem, social media platforms need to consider integrating tools and/or mechanisms that can help in the detection and prevention of such incidents. Therefore, to have a safer and more constructive social environment, it is necessary to design a smart network or an online patrol that will prohibit such behavior by monitoring and filtering the obscene, hateful, and improper content from social media posts.

4. DEVELOPED METHODOLOGY

1. First tweets are extracted with a tracking word, the output extracted is in Jason structure it contains data like user id, in_reply, is_quote_status, language etc.
2. After the tweets are collected the tweets are processed as they have incomplete information(inconsistent language), similar tweet up to 80% are removed for the data analysis.
3. During the processing the fields text and user description fields are merged, then the data is converted into the lowercase, the urls are removed also the non-alphanumeric characters. Thus, the clean-processed data required for analysis is obtained.
4. Using textblob library present in python the sentiment analysis (1,2,0) (Positive, negative, neutral) is done in the tweet.
5. A dictionary defines and loaded with the words offensive words and a dataset is loaded into the script for the further application of algorithms.
6. Another field taken into consideration is average offensiveness weight and percentages, this is defined as a function in the script.
7. Using sklearn in python with 5 different algorithms like linear gaussian NB is applied.
8. A training set is developed using machine learning which is then used to analyze the tweets and later in case of successive iterations.
9. After the analysis, the output is saved in a .csv file.

5.ANALYSIS AND CONCLUSION

after brief data gathering and processing we formed the text cloud which is a visual representation of text data, typically used to depict keyword metadata or to visualize free form text.ex.below text cloud is of metadata related to Syria.

figure1: text cloud formed by the tweets of keyword Syria.

figure 2: Bar chart of keyword Syria after analysis

and the **bar graph** represents the collected tweets using different algorithms like Logistic Regression, Linear Discriminant Analysis, KNeighboursClassifiers, Decision Tree Classifier, GaussianNB respectively.

due to increase in the internet slang and various acronyms for offensive words, some offensive weighted words are used as normal internet language which is still undetected by above algorithm.

since the AI is coming with better analysis algorithm, if we integrate the analysis with better machine learning

Future work:

With the Advent of machine learning and artificial intelligence, it is becoming easier to analysis text linguist. This help in an analysis of persons and their personalities in the virtual and cyberspace as well. Cybercrimes are in increase so does the method to prevent them has to become smarter. One of the crimes is cyberbullying and with proper steps, it can be prevented. It is a crime which affects the psychology of the person and can cause depression. This needs to be taken seriously and steps & policies need to be defined.

As seen from all the algorithm gaussianNB detection had a less false positive. Some offensive words which are taken into consideration in some parts are used as a causal internet lingo this makes it more complex to detect the correct results.

The most important step is making the users aware of the policies, laws related to such crimes.

6.REFERENCES

[1].Collaborative Detection Of Cyberbullying Behavior In Twitter Data by Amrita Mangaonkar

[2].Cyberbullying Among School-Aged Adolescents and Teens: A Policy Review and Recommendations for Georgia

[3.]Sentiment Informed Cyberbullying Detection in Social Media by Harsh Dani

[4] [Patchin.J.W., 2014] What is Cyberbullying? Retrieved from http://cyberbullying.org: http://cyberbullying.org/what-is-cyberbullying

[5   ]
Napolitano, Susan. (2011). Risk Factors for and Outcomes of Bullying and   Victimization. Education Psychology Papers and Publications.   Retrieved from http://digitalcommons.unl.edu/edpsychpapers/132/.

[6] Worried about your child's phone or tablet usage?. (2017). Retrieved from http://www.phonesheriff.com/: http://www.phonesheriff.com/