Detecting Early Signs of Cyberbullying in Social Media

Under the Guidance of Dr. Rishabh Kaushal
Department of Information Technology

Presented By : Neha Dubey(01602192021)
M.Tech in Cyber Security

Introduction

The Term CyberBullying is used to describe bullying taking place on the internet(through mobile phones, laptop etc.)

Bullying is generally used as an aggressive or intentional act or behaviour carried out by group or an individual repeatedly and over time against a victim who can not easly defend him or hers

Normal Bully Vs Cyber-Bully

Why it is important?

⅓ has been victim and nearly half of suicide case amoung 10-14 year olds are due to bullying

The American Academy of Pediatrics reports that children in third through fifth grades that own cell phones are more likely to be victims of cyberbullying. “Parents often cite the benefits of giving their child a cell phone, but our research suggests that giving young children these devices may have unforeseen risks as well,” said Elizabeth K. Englander, Ph.D., a professor of psychology at Bridgewater State University.

Why Children Do CyberBully?

- Revenge
- Frustration
- Entertainment
- Anger
- Others

Research Paper -

Nowadays, the amount of users' activities on online social media is growing dramatically. These online environments provide excellent opportunities for communication and knowledge sharing. However, some people misuse them to harass and bully others online, a phenomenon called cyberbullying. Due to its harmful effects on people, especially youth, it is imperative to detect cyberbullying as early as possible before it causes irreparable damages to victims. Most of the relevant available resources are not explicitly designed to detect cyberbullying, but related content, such as hate speech and abusive language. In this paper, we propose a new approach to create a corpus suited for cyberbullying detection. We also investigate the possibility of designing a framework to monitor the streams of users' online messages and detects the signs of cyberbullying as early as possible.

Keywords: Cyberbullying Detection, Text Mining, Early Text Categorization

Preparation of Datasets

Several works have been done towards finding cyberbullying traces on social media by detecting online hateful and aggressive comments. Still, most of these efforts are focused on offline settings and only detect the event after it took place. Therefore, none of these methods can be used for prevention.

In this research, we aim to detect early signs of cyberbullying using *as few textual evidence as possible* by providing *timely* predictions. The main contributions of this work are listed as follows:

- A new methodology for creating a cyberbullying corpus and the first dataset suited for the task of early cyberbullying prediction.

Data Collection

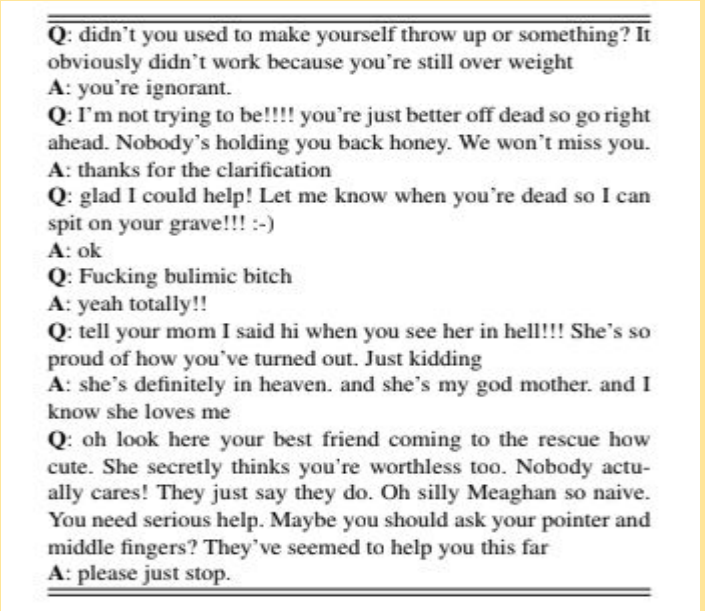
Abusive language detection can be considered as the initial step towards finding cyberbullying incidents. Cyberbullying happens when the victim receives several offensive messages repeatedly. Therefore, at least parts of the users' conversations should be monitored to detect such episodes.

Author collect data from ask.fm.4

Users can ask questions anonymously or with their identity attached. The platform also allows paid members and top users to start private conversations. Others can converse in public with or without their identity revealed.

Ask.fm allows users to post answers on their profile. These can be in text, video, or picture format. The answers you post can be liked by others or reacted to with a fire emoji. Currently, the mobile app is more interactive than the browser version.

We collect a large amount of ask.fm data, including the full history of question-answer pairs for 3K users. The question field includes a question/comment posted by the other users, and the answer field consists of the reply to that question/comment provided by the owner of t



Q: didn't you used to make yourself throw up or something? It obviously didn't work because you're still over weight
A: you're ignorant.
Q: I'm not trying to be!!!! you're just better off dead so go right ahead. Nobody's holding you back honey. We won't miss you.
A: thanks for the clarification
Q: glad I could help! Let me know when you're dead so I can spit on your grave!!! :-)
A: ok
Q: Fucking bulimic bitch
A: yeah totally!!
Q: tell your mom I said hi when you see her in hell!!! She's so proud of how you've turned out. Just kidding
A: she's definitely in heaven. and she's my god mother. and I know she loves me
Q: oh look here your best friend coming to the rescue how cute. She secretly thinks you're worthless too. Nobody actually cares! They just say they do. Oh silly Meaghan so naive. You need serious help. Maybe you should ask your pointer and middle fingers? They've seemed to help you this far
A: please just stop.

We utilize the ask.fm corpus proposed in the same work for training the model and label each row of our data automatically. To make the cyberbullying instances, we create a fixed-length sliding window and move it through the whole history of question-answer pairs per user. For each window sample, we calculate the ratio of offensive questions/comments that the user received inside the window. If it is greater than a pre-defined threshold, we consider the window as a potential cyberbullying event.

We divide all training and test examples to 10 different chunks to make the corpus suitable for early text classification. For every instance, each chunk contains 10% of all thequestion-answer pairs for that user.

Corpus creation:

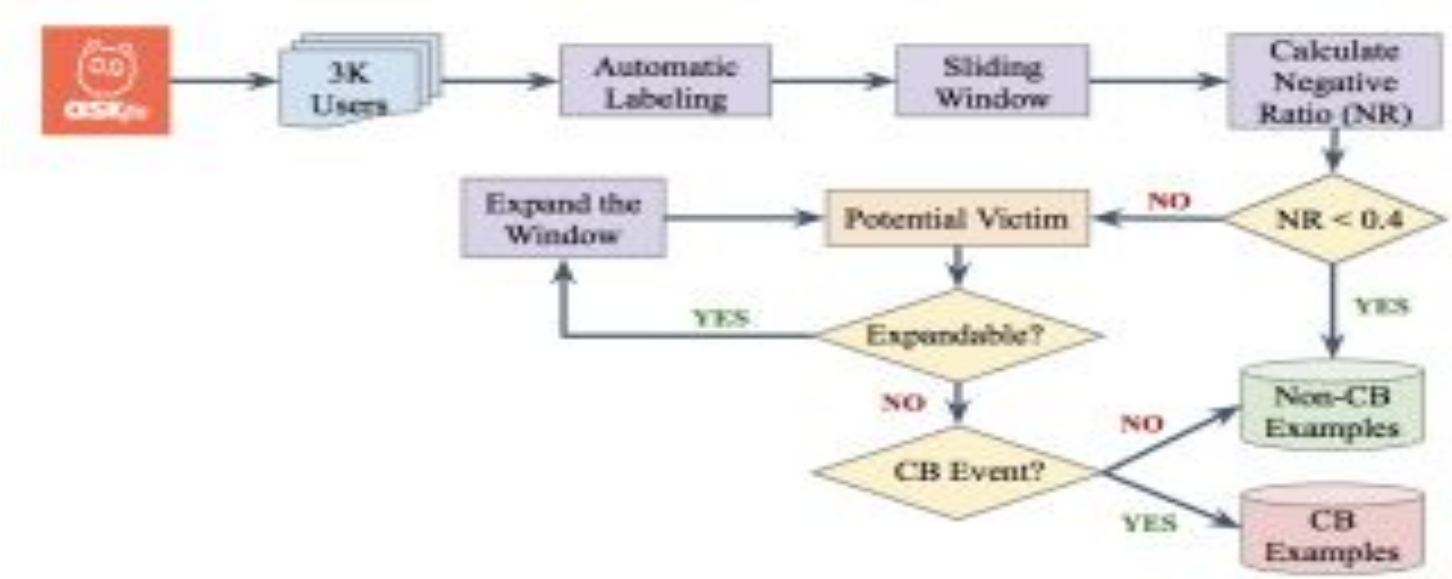


Figure 1: Overall process of building the new corpus.