TOXIC COMMENT CLASSIFICATION

PROJECT/RESEARCH

GROUP-B

PH21AIML-106

|  |  |  |
| --- | --- | --- |
| S.No. | NAMES | UID |
| 1. | Kaustubh | 21BCS6306 |
| 2. | Shashank | 21BCS |
| 3. | Arnav | 21BCS |
| 4. | Karthik | 21BCS |
| 5. | Anshuman | 21BCS |

TABLE OF CONTENTS

1. INTRODUCTION
2. PROBLEM STATEMENT
3. KEY-FEATURES/BENEFITS AND FUTURE SCOPE
4. EXECUTION
5. REFERENCES

INTRODUCTION

Over the years, social media and social networking use have been increasing exponentially due to an upsurge in the use of the internet. Flood of information arises from online conversation in a daily basis as people are able to discuss, express themselves and air their opinion via these platforms. While this situation is highly productive and could contribute significantly to the quality of human life, it could also be destructive and enormously dangerous. While discussion or a conversation is opened, it is quite obvious that debates may arise due to differences in opinion. But often these debates take a dirty side and may result in fights over the social media during which offensive language termed as toxic comments may be used from one side. These toxic comments may be threatening, obscene, insulting or identity- based hatred. So, these clearly pose the threat of abuse and harassment online. Consequently, some people stop giving their opinions or give up seeking different opinions which result in unhealthy and unfair discussion. As a result, different platforms and communities find it very difficult to facilitate fair conversation and are often forced to either limit user comments or get dissolved by shutting down user comments completely.

PROBLEM STATEMENT

Usually, these toxic comments are streamlined either by moderators for chatrooms or by simply banning the use of particular words (censored on the use) by the platforms. But these traditional methods fail to work in cases of floods of such comments or due to small changes in words with replacement of a few letters. And thus slipping through into the comment sections and ultimately success in attacking the targeted audience.

KEY FEATURES/BENEFITS AND FUTURE SCOPE

We hope to implement our project in chatrooms/comment sections which could lead us

* To distinguish these toxic comments from normal comments
* To create a healthy chatspace for people with fair conversation opportunities
* To ultimately fight off the people using the social medias to spread hatred and asserting their personal opinions regardless of the people which are being attacked directly/indirectly
* To create a safe space for children who often end up being hampered due unsafe environments in chatrooms/comment sections.

This project would directly apply some of the already existing models to be able to implement in chatrooms and comment sections.

TOOLS USED

-GOOGLE COLAB

-VSCODE

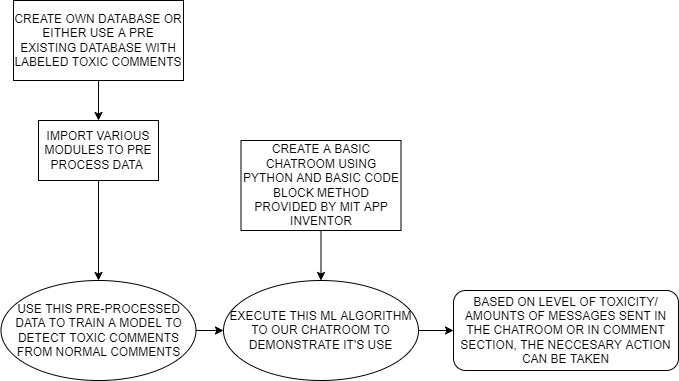
-NLP WITH PYTHON USING NLTK

-BASIC MODULES(NLTK, Discord, Socket, Select)

-SIDE MODULES(collections, string, pandas, colorama, matplotlib)

-MIT App Inventor

EXECUTION



REFERENCES

* <https://www.rsisinternational.org/journals/ijrias/DigitalLibrary/Vol.4&Issue11/142-147.pdf>
* <https://www.youtube.com/watch?v=eObouMO2qSE&list=PLQVvvaa0QuDf2JswnfiGkliBInZnIC4HL&index=19>
* https://towardsdatascience.com/how-vicious-do-you-think-your-social-media-comments-are-204128141eab