CSD 350: Natural Language Processing

HateDetect

A hate detection system

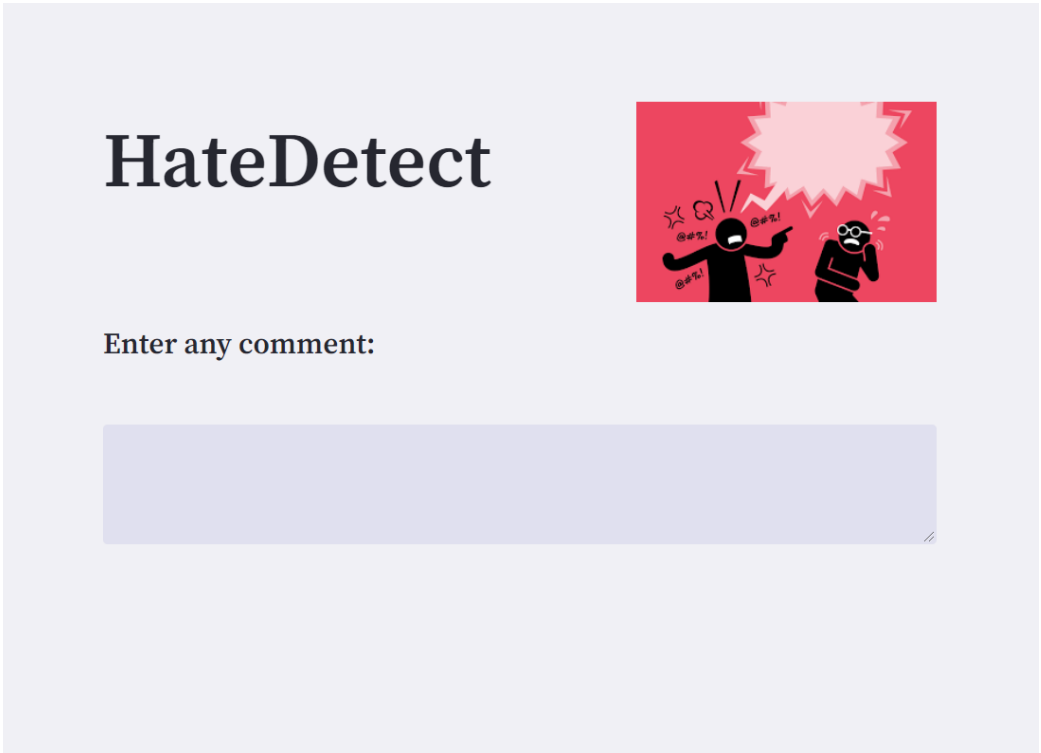
Final Project Report

Submitted by:
Aditya Srivastava
1910110034

Contents

What is HateDetect?

- What is Hate Speech?
- Difficult to create as no legal definition
- Built to detect hate speech on different platforms.



Motivation behind HateDetect

- Due to the societal concern and how widespread hate speech is becoming on the Internet, there is strong motivation to study the automatic detection of hate speech. By automating its detection, the spread of hateful content can be reduced.
- By creating this application various social media platforms will become a lot more user-friendly by providing the users a way safer environment by preventing cyberbullying, harassment, teasing, etc.

Executing the Program

Pre-Requisites

1. An IDE that supports Python.
2. A web browser
3. Before execution run this command in the terminal (ensure you are in the correct folder)
"**pip install streamlit**"

How to Start?

Hate Speech Detection->
Hate_Speech_Detection.py->Open terminal->
Type "streamlit run Hate_Speech_Detection.py"

Dataset Used

1. Dataset has been taken from **Kaggle**.
2. Due to the nature of the study, it's important to note that this dataset contains text that is considered hateful because of being racist, sexist or homophobic, etc.
3. The dataset has **7 columns** in total
    - index
    - count
    - hate_speech
    - offensive_language
    - neither
    - class
    - comment

Libraries

The following libraries have been imported:

1. streamlit **as** st
2. Image **from** PIL
3. string
4. stopwords **from** nltk.corpus
5. nltk
6. re
7. pr **from** nltk.util
8. pandas **as** pd
9. numpy **as** np
10. CountVectorizer **from** sklearn.feature_extraction.text
11. train_test_split **from** sklearn.model_selection
12. DecisionTreeClassifier **from** sklearn.tree

Analysis of Program

1. First import all the libraries that will be required in the model.
2. Then we read the dataset and add a new column 'labels' by mapping it with the 'class' column of the dataset. The three labels are:
    a. Normal Language
    b. Hate Language Detected
    c. No Hate
3. Now we choose the 'comment' and 'labels' columns for the task of training our model.
4. We create a function for cleaning our text from any unnecessary data that is not necessary for the analysis.
5. We split our dataset for training and testing such that a **33% proportion** of the dataset is within them. To get the same data for analysis we have set the random seed to be **42**.
6. Now we create a decision tree classifier that gives an accuracy score for our test data.
7. Finally, we run our "hate_speech_detection()" function which opens the local website in the browser (through the help of streamlit), you add your comment, and the function analyses it and shows on the website whether it is a hate comment or not.

HateDetect

Enter any comment:

Hey There!

['Result: No Hate']

HateDetect

Enter any comment:

I will kill everyone.

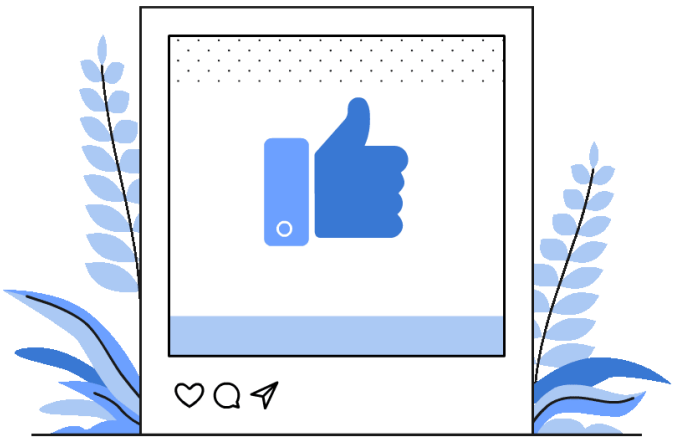
['Result: Hate Language Detected']

Limitations

- Words can be obfuscated.
- Hate Speech is different for every individual, expressions that are not inherently hateful, can be so in some different context.

Future Aspect

- As part of the future scope, HateDetect can be expanded by taking a bigger dataset to combat the hate.
- It can be incorporated with forums to make the applications community-friendly.



Key Take Aways

- Creating this project was a great learning experience. It helped me explore the different workspace altogether by exposing me to new platforms like Kaggle, use of datasets, and above all I got to apply the learnings of my course.
- We live in a world of Social Media and through this project, I could understand the intricacies and consequences of the hate language used online.
- I am happy that I got to create something which promotes a healthy safe environment for users online.

Acknowledgement

I would like to extend a token of gratitude to **Prof. Ketan Bajaj** for his continued guidance and constructive feedback given towards enhancing the quality of the project.

His views helped me to accomplish the objectives of the project in a well-defined manner.

*****

THANK YOU.