

CYBERBULLYING TWEET ANALYSIS

Div:-AI -B EDI_Group No:- 2

Day:-Friday

Date:-16|12|22

Presented by

| | |
|-----------------------|----------|
| 45 - Samved Patil | 12110152 |
| 48 - Siddhi Patil | 12111408 |
| 55 - Vaishnavi Petkar | 12110717 |
| 56 - Mrunmayee Phadke | 12110650 |

Project Guide

Prof. Puja Cholke

Department of Multidisciplinary Engineering

VISHWAKARMA INSTITUTE OF TECHNOLOGY, PUNE

Contents

1 Introduction

2 Literature Survey

3 Problem Statment

4 Objectives

5 System Architecture

7 System Requirments

8 Demo of Our Project

9 Advantages and challenges

10 Conclusion and Future Scope

12 References

Introduction

- People spend more time on social media, that keeps them constantly connected to other people, giving room for cyberbullying and creating the need of cybersecurity.
- 85% of Indian children are cyberbullied making it global highest.



- Cyberstalking, impersonation, and harassment are types of cyberbullying.
- Cyberbullies can communicate with their victims through text messaging, social networking websites, and instant messaging over the Internet.

Literature Survey

| Title of Paper | Authors | Methodology | Demerits | Publication Year |
|--|---|--|---|---|
| <ul style="list-style-type: none">• AI-Enabled Cyberbullying-Free Online Social Networks in Smart Cities• Cyber Bullying Detection using NLP and Text Analytics | <ul style="list-style-type: none">• Abdulsamad Al-Marghilani• Yeo Khang Hsien, Zailan Arabee Abdul Salam, Vinothini Kasinathan | <ul style="list-style-type: none">• AI-enabled cyberbullying-free OSN (AICBF-ONS) technique in smart cities.• NLP and python based application removes inappropriate tweet and informs guardian about the case via email. | <ul style="list-style-type: none">• Performance of the AICBF-ONS technique could be extended to the design of outlier detection and data clustering approaches in big data environment• Domain-specific, ambiguity | <ul style="list-style-type: none">• 2022• 2022 |

| Title of Paper | Authors | Methodology | Demerits | Publication Year |
|--|--|---|--|---|
| <ul style="list-style-type: none">• An NLP-Assisted Bayesian Time Series Analysis for Prevalence of Twitter cyberbullying during COVID-19 pandemic• Analysing Cyberbullying using Natural Language Processing by Understanding Jargon in Social Media | <ul style="list-style-type: none">• Christopher Perez,Sayar Karmarkar• Bhumika Bhatia, Anuj Verma, Anjum, Rahul Katarya | <ul style="list-style-type: none">• A Pre-trained NLP Model for Pre-processing Data, Sub setting Truly Offensive or Hateful Tweets, Visual Analysis on the Raw and Filtered Counts• This paper explores binary classification on multiple datasets. It uses multiple models suchas Bi-LSTM, GloVe, state-of-the-art models like BERT and unique pre-processing technique | <ul style="list-style-type: none">• It's necessary to develop a model with a new set of data without using a pre-trained model, It can take weeks to achieve a good performance depending on the amount of data.• This model detects various types of cyberbullying but does not classify them. | <ul style="list-style-type: none">• 2022• 2021 |

| Title of Paper | Authors | Methodology | Demerits | Publication Year |
|---|--|--|---|---|
| <ul style="list-style-type: none">• Crime Detection and Analysis from Social Media Messages Using Machine Learning and Natural Language Processing Technique• Analysis of fake news comments in social media | <ul style="list-style-type: none">• Xolani Lombo, Oyelade Olaide, Absalom El-Shamir Ezugwu• Yulia M. Kuznetsova, Maxim Stankevich | <ul style="list-style-type: none">• The system can detect various crimes using Natural language processing algorithms, SVM and classifiers to classify texts.• The paper trained model on a huge dataset that was non-biased. | <ul style="list-style-type: none">• The accuracy of the models could be increased. <p>-----</p> | <ul style="list-style-type: none">• 2022• 2022 |

| Title of Paper | Authors | Methodology | Demerits | Publication Year |
|---|--|--|---|---|
| <ul style="list-style-type: none">• Cyberbullying Detection in Social media using Machine Learning• Unsupervised Cyber Bullying Detection in Social Networks | <ul style="list-style-type: none">• Aditya Desai, Shashank Kalaskar, Omkar Kumbhar, Rashmi Dhumal• Michele Di Capua, Emanuel Di Nardo | <ul style="list-style-type: none">• Proposed model based to be considered while detecting cyberbullying and implement bidirectional deep learning model - BERT.• Propose solution for automatic detection of bully traces over a social network, using techniques derived from NLP. | <ul style="list-style-type: none">• Models have not considered all the necessary features that can be used to identify or classify a statement or post as bullying• Clustering efficiency of documents containing bully traces, built upon semantic and syntactic features of textual sentences. | <ul style="list-style-type: none">• 2021• 2020 |

| Title of Paper | Authors | Methodology | Demerits | Publication Year |
|---|---|---|---|---|
| <ul style="list-style-type: none">• Cyberbullying detection using fairness constraints.• Linking textual and contextual features for intelligent cyberbullying detection in social media | <ul style="list-style-type: none">• Oguzhan Gencoglu• Nabi Rezvani, Amin Beheshti, Alireza Tablordingbar | <ul style="list-style-type: none">• A cyberbully detection model where instead of manipulating dataset. They use fairness constraints of multiple rules to make a model that is context-based.• This paper provides a pipeline for cyberbully detection that is<ul style="list-style-type: none">i. Extraction of featuresii. Contextualizing the features using crowdsourced feedback in a loopiii. Combining features using the fully connected network to identify and build potential useful features. | <ul style="list-style-type: none">• The datasets used are common ones. and the Twitter dataset has the lowest accuracy of 11.33%• This model appears to be specifically designed for Twitter and Instagram. They haven't thought about using any other social media sites. | <ul style="list-style-type: none">• 2020• 2020 |



Problem Statement

CYBERSECURITY AND CYBERBULLYING ANALYSIS IN SOCIAL MEDIA TO PREVENT A HOSTILE ENVIRONMENT THAT ERODES ENGAGEMENT AND COLLABORATION OF PEOPLE AND TO MITIGATE THE HEAVY SOCIAL PROBLEM OF CYBER BULLYING.

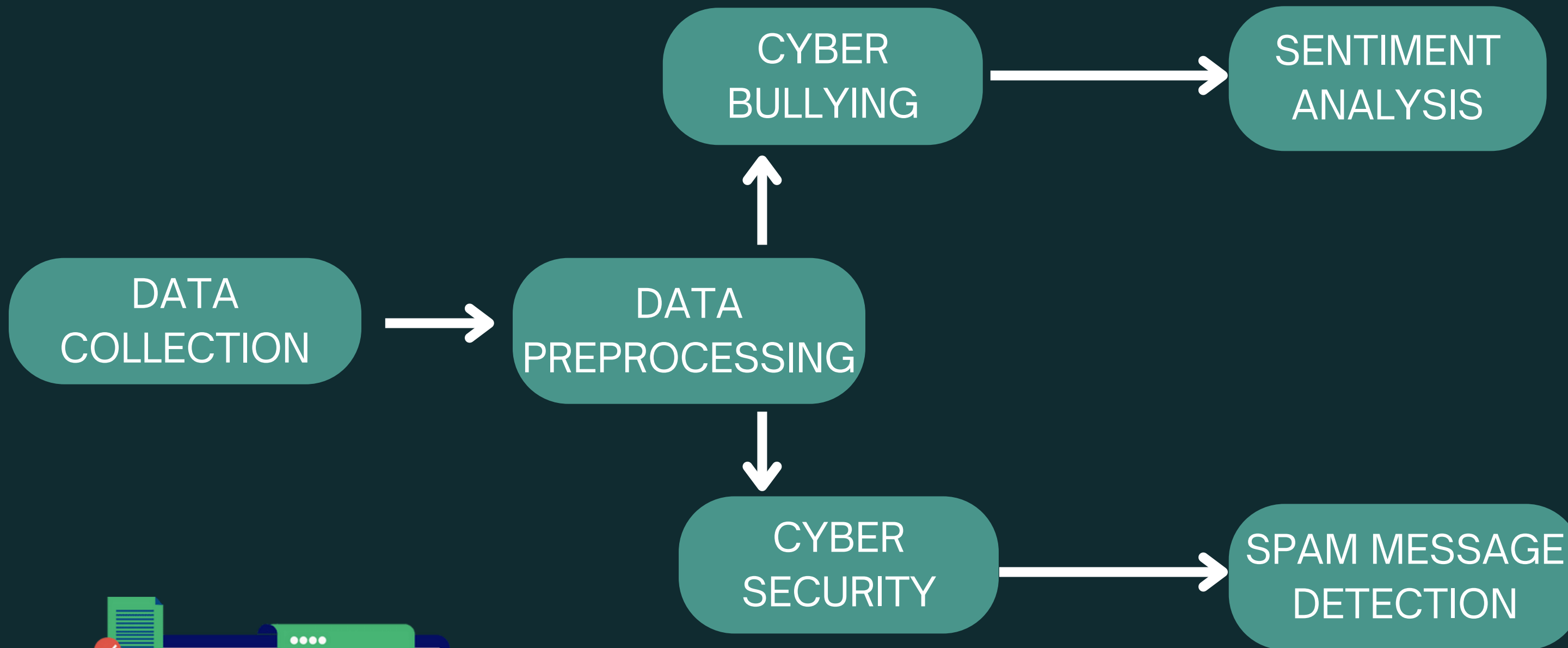


Objectives

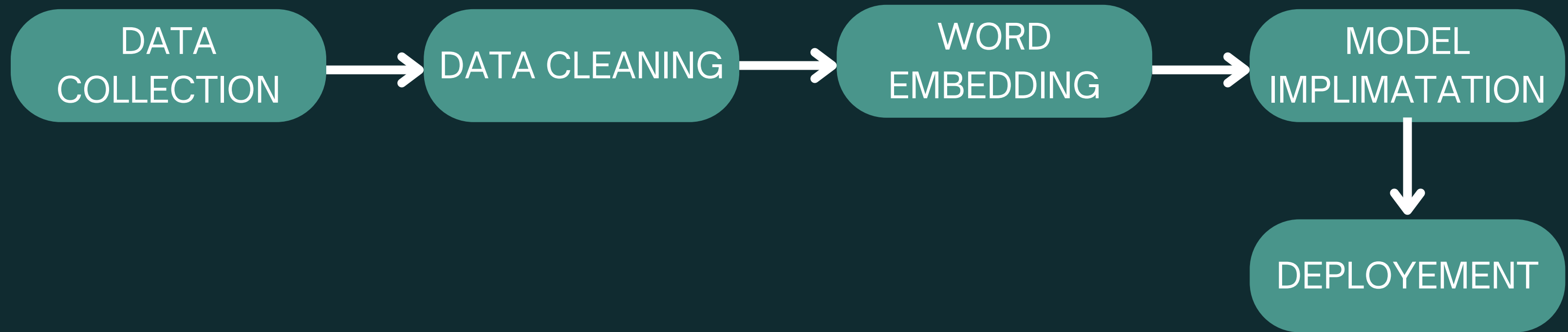
- To construct literature survey.
- To find unbiased, large and suitable datasets.
- To study appropriate data preprocessing techniques.
- To research various algorithms for the problem statement.
- To detect :
 - Cyberbullying attack vectors
 - Cybersecure elements



System Architecture

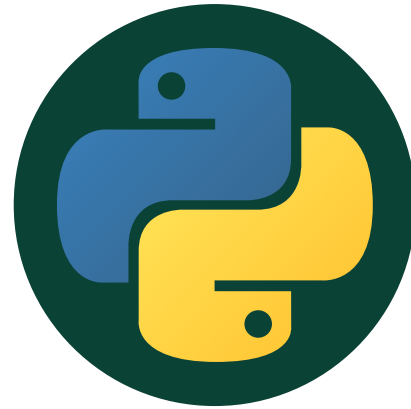


System Architecture of Cyberbullying



System Requirements

Tools:

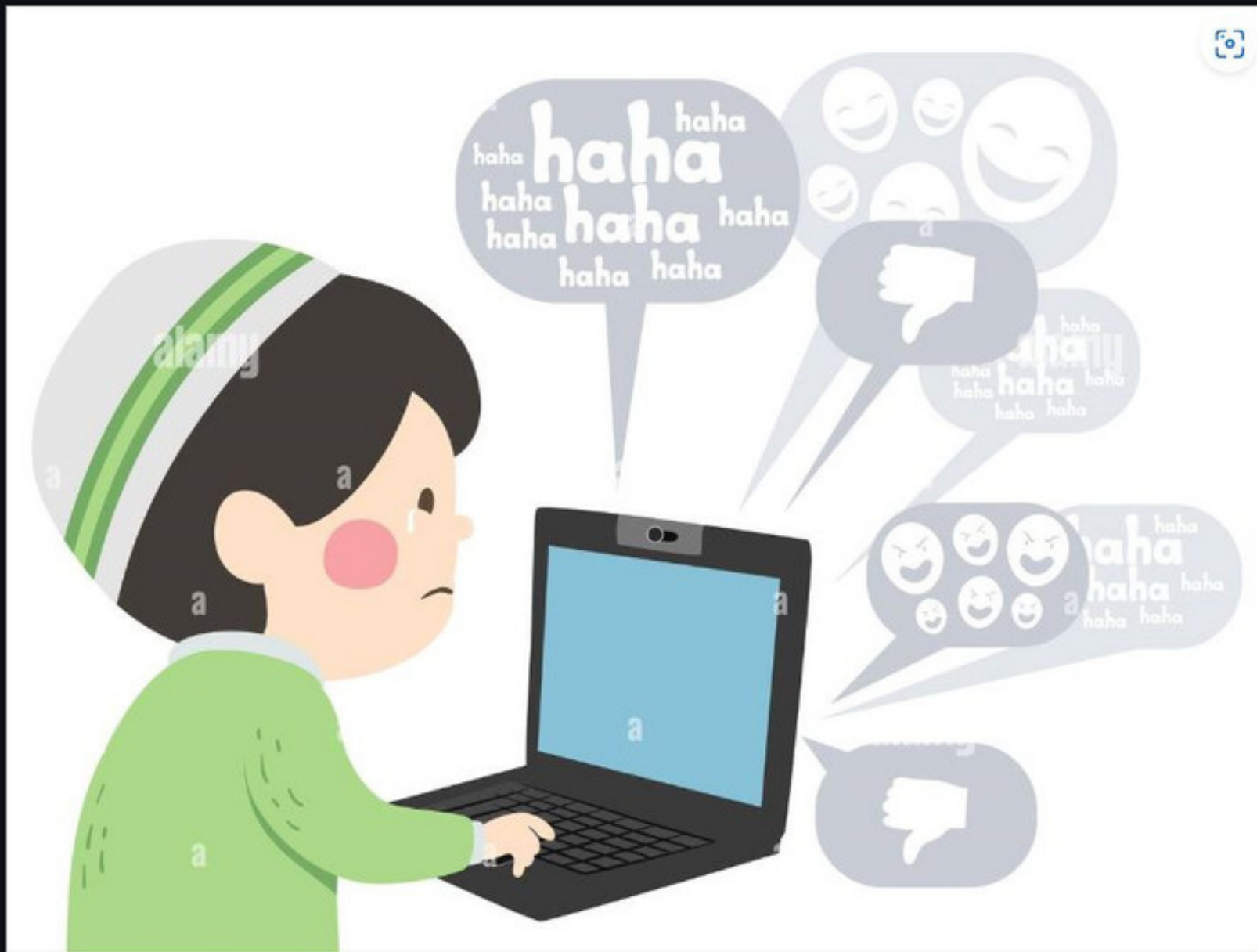


Libraries:



Demo of Our Project

Prediction



Religion Cyberbullying!

Cyberbullying Tweet Analysis App

This app predicts the nature of the tweet into the following Categories.

- Bullying
- Non-Bullying
- Age
- Ethnicity
- Gender
- Not Cyberbullying
- Other Cyberbullying
- Religion

Enter Tweet

Tweet Input

But for u its Hinduphobia isnt it? When kashmiri pandits get killed, when a hindu girl gets raped by islamists, when radical islamic terrorism kill people in the world,u still keep quiet as if nothing is happening;but jump on when some1 says anything against islam!! #Hinduphobic

Entered Tweet text

But for u its Hinduphobia isnt it? When kashmiri pandits get killed, when a hindu girl gets raped by islamists, when radical islamic terrorism kill people in the world,u still keep quiet as if nothing is happening;but jump on when some1 says anything against islam!! #Hinduphobic

Demo of Our Project

```
[46]: # Model
from sklearn.svm import SVC
svm_model_linear = SVC(kernel='linear', C=1).fit(X_train, y_train)
svm_predictions = svm_model_linear.predict(X_test)
accuracy = svm_model_linear.score(X_test, y_test)
print(accuracy)
```

0.8212674323215751

```
[49]: print(classification_report(y_test, svm_predictions))
print(confusion_matrix(y_test, svm_predictions))
print('Accuracy', accuracy_score(y_test, svm_predictions))
```

| | precision | recall | f1-score | support |
|---------------------|-----------|--------|----------|---------|
| Bullying | 0.53 | 0.37 | 0.43 | 90 |
| Non-Bullying | 0.60 | 0.22 | 0.32 | 123 |
| age | 0.92 | 0.99 | 0.95 | 1574 |
| ethnicity | 0.95 | 0.99 | 0.97 | 1639 |
| gender | 0.91 | 0.87 | 0.89 | 1631 |
| not_cyberbullying | 0.62 | 0.49 | 0.55 | 1588 |
| other_cyberbullying | 0.59 | 0.69 | 0.63 | 1521 |
| religion | 0.93 | 0.96 | 0.95 | 1586 |
| accuracy | | | 0.82 | 9752 |
| macro avg | 0.76 | 0.70 | 0.71 | 9752 |
| weighted avg | 0.82 | 0.82 | 0.82 | 9752 |

```
[[ 33  11   1   7   8   4  23   3]
 [ 15  27   2  11   8  12  46   2]
 [   0   0 1554   1   2   6  11   0]
 [   1   0   1 1621   4   4   8   0]
 [   5   1   8  141  89  87   8]
 [   3   5  99  20  58 781 551  71]
 [   5   1  29  19  63 332 1047  25]
 [   0   0   3   9   5  32  10 1527]]
```

Accuracy 0.8212674323215751

```
[0]: print(classification_report(y_test, y_pred1))
print(confusion_matrix(y_test, y_pred1))
print('Accuracy', accuracy_score(y_test, y_pred1))
```

| | precision | recall | f1-score | support |
|---------------------|-----------|--------|----------|---------|
| Bullying | 0.36 | 0.31 | 0.33 | 90 |
| Non-Bullying | 0.38 | 0.24 | 0.30 | 123 |
| age | 0.98 | 0.97 | 0.97 | 1574 |
| ethnicity | 0.97 | 0.96 | 0.97 | 1639 |
| gender | 0.83 | 0.81 | 0.82 | 1631 |
| not_cyberbullying | 0.46 | 0.45 | 0.45 | 1588 |
| other_cyberbullying | 0.46 | 0.54 | 0.50 | 1521 |
| religion | 0.94 | 0.90 | 0.92 | 1586 |
| accuracy | | | 0.76 | 9752 |
| macro avg | 0.67 | 0.65 | 0.66 | 9752 |
| weighted avg | 0.77 | 0.76 | 0.76 | 9752 |

```
[[ 28  10   1   4   9   8  27   3]
 [ 23  30   1   4   3  24  32   6]
 [   0   0 1522   1   2  29  20   0]
 [   5   5   1 1571  11  16  23   7]
 [   3   7   2   6 1321 125 159   8]
 [   6   7  19  14 120 709 660  53]
 [   7   9   9   7 110 548 823   8]
 [   6  10   2   9  20  78  39 1422]]
```


Advantages

- Helps identify various forms of cyberbullying and threats to cybersecurity that can be used in different collaborations.
- Helps create a safe environment for people of all age, gender and race.



Challenges

- The access restrictions on high-quality data limit the applicability of state-of-the-art techniques.
- Compiling various types of cyberbullying ways under one general head.



Conclusion

Cybersecurity and tackling cyberbullying is the need of the hour.

Future Scope

COLLABORATION

We can collaborate with various social media platforms like twitter or reddit and integrate our algorithm with theirs. We can also collaborate with the government to reduce cyber crime and harrasement.

EXTENSION

We can create a chrome extension. Using web scrapping we can collect the data on user's screen and implement our model on it to alert the user.

AWARENESS AND REPORTS

We can collect reports and statistics around the current scenario of cybersecurity and cyberbullying in social media. We can create awareness and take measures in required areas based on these statistics.



References

- Oguzhan Gencoglu, Cyberbullying Detection with Fairness Constraints, IEEE Xplore,2020
- Nabi Rezvani, Amin Beheshti, Alireza Tablordingbar, Linking textual and contextual features for intelligent cyberbullying detection in social media, Association for Computing Machinery New York, NY, United States,2021 [1]
- Aditya Desai, Shashank Kalaskar, Omkar Kumbhar, Rashmi Dhumal, Cyberbullying Detection in Social media using Machine Learning, ITM web of conferences,2021 [2]
- Michele Di Capua, Emanuel Di Nardo, Unsupervised Cyber Bullying Detection in Social Networks, 2016 23rd International Conference on Pattern Recognition (ICPR),2020[3]
- Yeo Khang Hsien, Zailan Arabee Abdul Salam, Vinothini Kasinathan, Cyber Bullying Detection using NLP and Text Analytics, IEEE International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE),2022[4]
- Christopher Perez, Sayar Karmarkar, An NLP-Assisted Bayesian Time Series Analysis for Prevalence of Twitter cyberbullying during COVID-19 pandemic,Cornell university,2022[5]
- Yulia M. Kuznetsova, Maxim Stankevich, Analysis of fake news comments in social media,Journal of Retailing and Consumer Services ,Volume 57, November 2020[6]

Thank You!

We are open to questions.

