Cyberbullying Detection

Using RNN and hybrid LSTM

Introduction

Cyberbullying is a serious issue that can have devastating effects on victims. Traditional methods of detection often fall short, especially when dealing with subtle forms of cyberbullying. Deep learning techniques, such as Recurrent Neural Networks (RNNs) and hybrid models combining RNNs and Convolutional Neural Networks (CNNs), offer a promising solution. These models can analyze text data, identify patterns, and accurately classify content as cyberbullying or non-cyberbullying. By leveraging the power of deep learning, we can create safer online environments and protect individuals from the harmful effects of cyberbullying.

Milestone 1

- Data Collection: web scraping
  Social media platforms used : Youtube
  Link - https://youtu.be/YgQy70_LPS4?si=aR3qko1M_niA8VAN
- Data Preprocessing
- Class labelling
- Text Cleaning - normalized text(convert to lower case, stemming)
- Removing contractions and punctuations
- Tokenization – Breaking text into tokens
- Stop words Removal(e.g: is,the,and,of etc.)
- Data Splitting(Divide the dataset into training, validation, and testing sets.)

By leveraging web scraping and advanced text preprocessing techniques, we can effectively collect and prepare a robust dataset for training our cyberbullying detection model.

Milestone 2

Implementing Traditional ML models:

- Logistic Regression
  - Text preprocessing-Convert text data into numerical features (e.g., TF-IDF, bag-of-words).
  - Model Training-Train a logistic regression model to classify text as cyberbullying or non-cyberbullying.
- Random Forest Classifier
  - Text preprocessing- similar to Logistic regression
  - Model Training- Train a random forest classifier to classify text.
- Performance
  - Evaluate model performance using metrics like accuracy, precision, recall, and F1-score.
  - Use confusion matrices to visualize classification results.

WEB SCRAPING

Web scraping is the process of using automated tools or scripts to extract specific data from websites. It involves sending requests to web servers, receiving HTML responses, parsing the content, and extracting information based on HTML tags or patterns. This data is stored in a structured format for easy analysis. Web scraping is widely used in fields like market research, price monitoring, news aggregation, and social media tracking.

- **Sending a Request:** The scraper first sends a request to the website's server for a specific webpage (usually an HTTP GET request).
- **Retrieving the HTML:** If the request is successful, the server responds with the HTML content of the webpage.
- **Parsing the HTML:** The scraper then processes the HTML to locate and extract specific information. It involves identifying particular HTML tags, classes, IDs, or other elements that contain the desired data.
- **Extracting and Storing Data**: Once the information is identified, it’s extracted, transformed as needed, and stored in a structured format like CSV, Excel, or a database for easy access and analysis.

Pre-processing:

Pre-processing cleans and structures raw data for analysis by removing irrelevant symbols, correcting errors, and handling missing values. It ensures consistency and reduces noise, which is essential for accurate results. Missing data can be imputed, flagged, or removed depending on its relevance. Pre-processing transforms messy data into a reliable, usable format, improving the quality and performance of machine learning models or analytical tools.

Text Normalization:

Text normalization standardizes text data for analysis in natural language processing. It includes converting text to lowercase, removing stop words (like "the" or "is") that add little value, stripping punctuation, and applying lemmatization to reduce words to their root forms. This process ensures consistency, reduces redundancy, and focuses on the most meaningful parts of the text, enabling more accurate analysis and machine learning results.

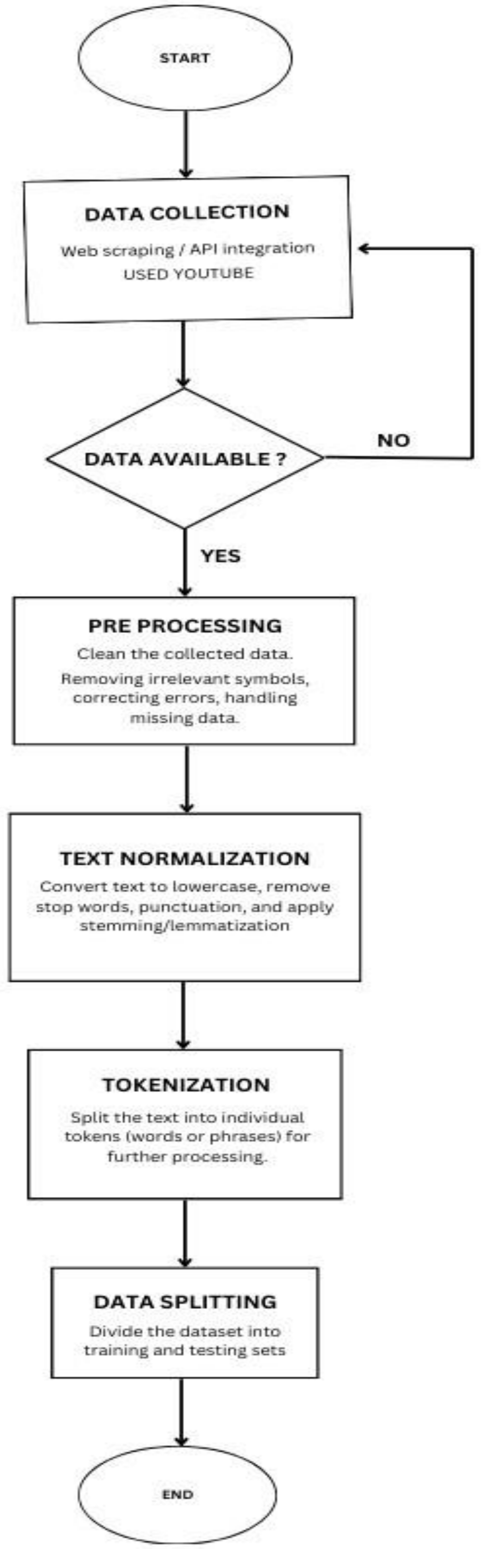
Tokenization and Text Conversion

- Tokenization converts text into smaller units, typically words or sub-words, to help the model understand the input text. This is necessary for processing text data in machine learning models.

- **Why Tokenization?** Cyberbullying detection requires analyzing the semantics and context of sentences. Breaking down text into tokens allows the model to handle each word as a meaningful unit.

1. Converts text to lowercase, making it easier to work with consistently.
2. Tokenization happens here indirectly: the text is split by whitespace (via split()) and joined back together, giving us lowercase tokens without changing word order.

Flow chart


START
DATA COLLECTION
Web scraping / API integration
USED YOUTUBE
DATA AVAILABLE ?
NO
YES
PRE PROCESSING
Clean the collected data.
Removing irrelevant symbols, correcting errors, handling missing data.
TEXT NORMALIZATION
Convert text to lowercase, remove stop words, punctuation, and apply stemming/lemmatization
TOKENIZATION
Split the text into individual tokens (words or phrases) for further processing.
DATA SPLITTING
Divide the dataset into training and testing sets
END

ACCURACY RATE

LOGISTIC REGRESSION:

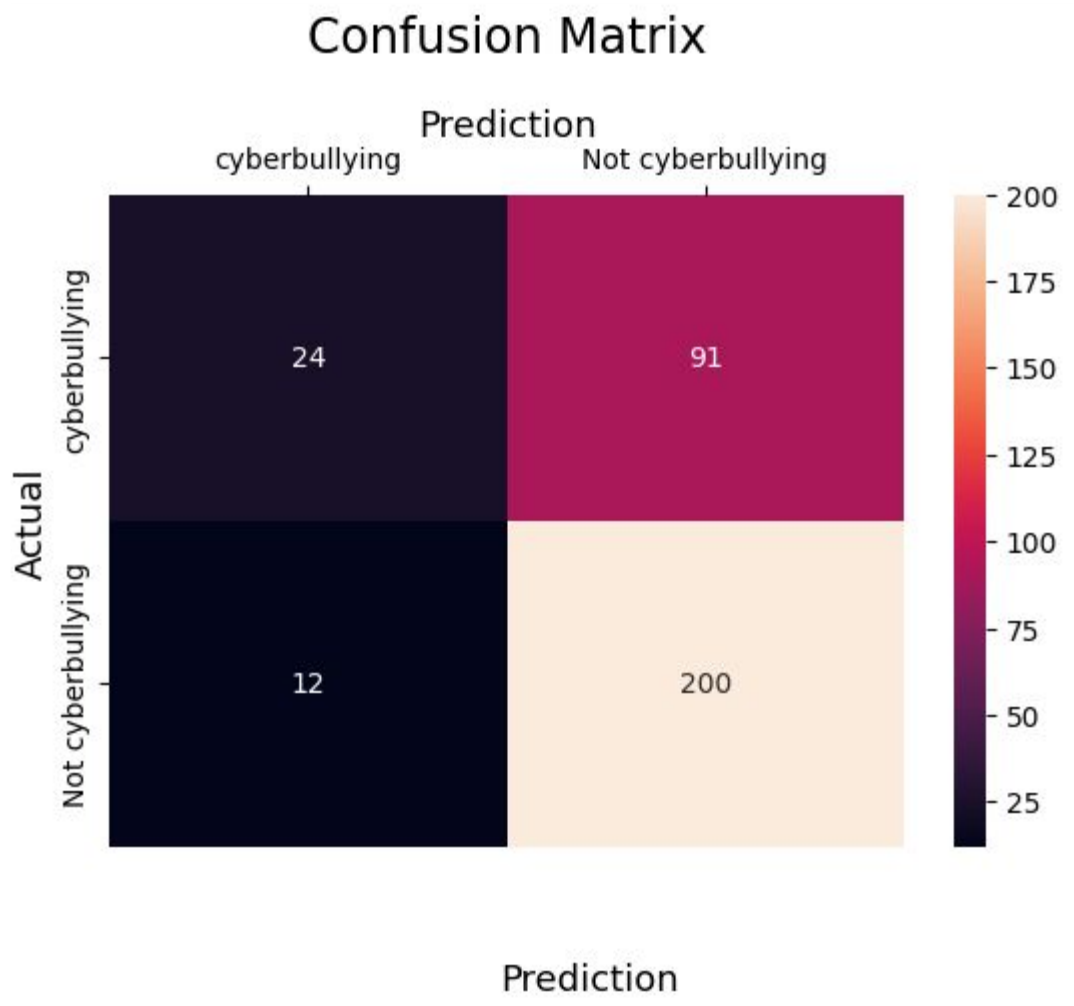
- For test size 0.2 : 70.70 %
- For test size 0.25 : 70.45 %
- For test size 0.3 : 69.36 %
- For test size 0.33 : 68.50

RANDOM FOREST REGRESSION

- For test size 0.2 : 68.68 %
- For test size 0.25 : 68.01 %
- For test size 0.3 : 70.03 %
- For test size 0.33 : 71.25 %

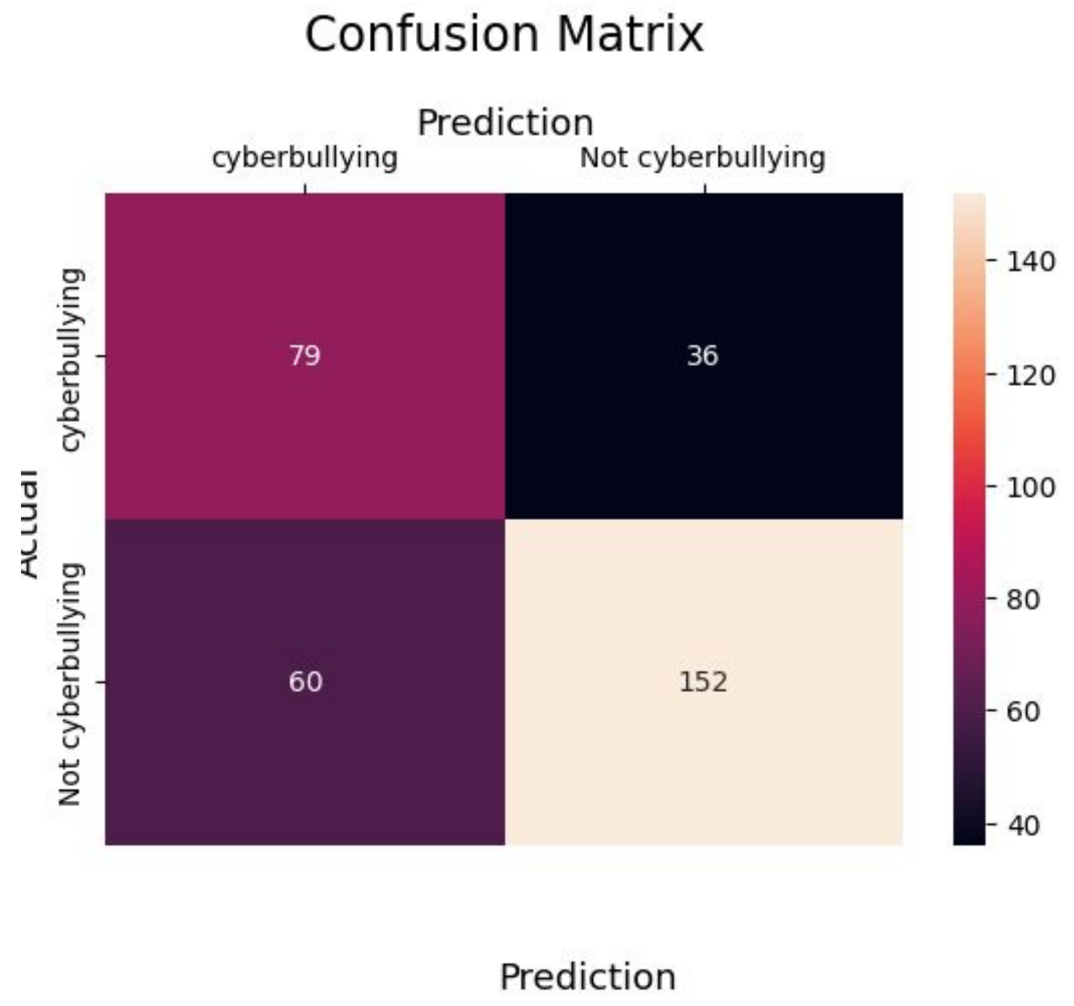
Confusion Matrix and Accuracy

Logistic Regression



Accuracy : 0.6850152905198776

Radom forest classifier



Accuracy : 0.7064220183486238

Conclusion

Cyberbullying, a pervasive issue in the digital age, poses significant threats to individuals, especially adolescents. Traditional methods often struggle to detect subtle forms of cyberbullying.

Deep learning, particularly Recurrent Neural Networks (RNNs) and hybrid models combining RNNs and Convolutional Neural Networks (CNNs), offer a promising solution. RNNs effectively capture long-term dependencies within text, while CNNs extract local features. These hybrid models can accurately identify cyberbullying patterns, even in complex and evolving online interactions.

By training these models on large datasets and fine-tuning their hyperparameters, we can achieve state-of-the-art performance in cyberbullying detection. This enables us to create safer online environments and protect individuals from the harmful effects of cyberbullying.

Team Members

- Janvi Singh
- Praneetha Sai
- Dona Krishna sai
- Ayush Srivastava
- Kalpesh Mahajan