International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 05 May 2024, pp: 840-845 www.ijaem.net ISSN: 2395-5252

Identifying Cyber bullying Using Unsupervised Learning

Date of Submission: 18-05-2024 Date of Acceptance: 28-05-2024

ABSTRACT—With the exponential increase of social media users, cyber bullying has been emerged as a form of bullying through electronic messages. Social networks provide a rich environment for bullies to uses these networks as vulnerable to attacks against victims. Given the consequences of cyberbullying on victims, it is necessary to find suitable actions to detect and prevent it. Machine learning can be helpful to detect language patterns of the bullies and hence can generate a model to automatically detect cyber-bullying actions. This project proposes a unsupervised machine learning approach for detecting and preventing cyberbullying. Several classifiers are used to train and recognize bullying actions.

Keywords-Cyberbullying detection and prevention, machine learning, LSTM algorithm

I.INTRODUCTION

In the current digital era, the extensive use of social media and online communication technologies has increased people's ability to harass, abuse, or threaten others, which has raised concerns about cyberbullying. Because cyberbullying episodes can negatively affect a person's mental health, sense of self, and general wellbeing, it is crucial to identify and stop them. Researchers and developers have concentrated on creating automated approaches to identify cyberbullying using machine learning and natural language processing to this issue. using machine learning and natural language processing to this issue. Using HMM and yolov8 bullying text and bullying hand gesture sign is detected

Cyberbullying is a complex and evolving issue that is influenced by various factors, including technological advancements, social media trends, and cultural shifts.In the interconnected and digital landscape of the 21st century, the rise of technology has brought unprecedented opportunities for communication and collaboration. However, it has also given birth to a darker phenomenon—cyberbullying. This pervasive issue transcends geographical boundaries, impacting individuals of all ages and backgrounds.

II.BACKGROUND

Some platforms use machine learning algorithms to analyze user behaviour and identify patterns associated with cyberbullying. AI models can help in automating the detection of cyberbullying by learning from vast amounts of data.

Text mining is valuable in cyber bullying detection as it helps analyze large volumes of text data to identify patterns and trends indicative of abusive behavior. Techniques like sentiment analysis, natural language processing, and machine learning can be employed to automatically detect and flag potential instances of cyber bullying in online communications. This enables quicker response and intervention to mitigate the impact on individuals affect.

This research proposes a new method for detecting and classifying cyberbullying activities on social media using fuzzy logic and genetic algorithms. The system analyze conversations for features like word types, statistics, and fuzzy rules to identify cyberbullying terms. This approach aims to overcome limitations of existing systems by capturing complex relationships and offering user control over rules. The research shows promise for improving cyberbullying detection and classification accuracy

Content filtering and moderation tools are essential components of online platforms to maintain a safe and positive user experience. These tools use various techniques to identify and manage potentially harmful or inappropriate content. Content filtering systems begin by collecting a vast amount of data, including text, images, videos, and user behaviour data from the platform. Machine learning algorithms are developed to analyze this data. Natural Language Processing (NLP) algorithms are common for text analysis, while image recognition and computer vision techniques are used for multimedia content.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 05 May 2024, pp: 840-845 www.ijaem.net ISSN: 2395-5252

III.SYSTEM DESIGN

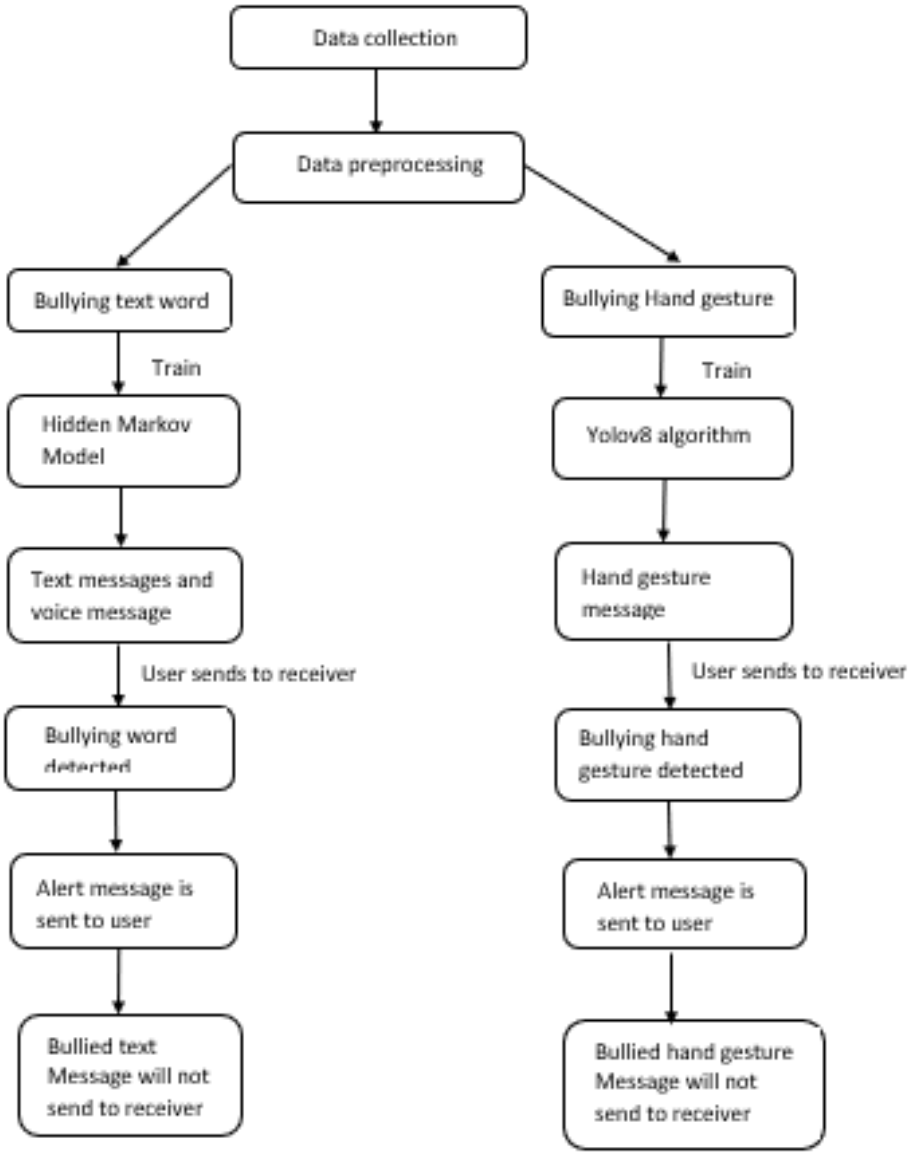

Fig.1.System Architecture

Designing a system for text cyberbullying and image detection using YOLOv8 algorithm involves several components**.** This project tackles cyberbullying through a multi-layered approach. It leverages machine learning to analyze text messages for bullying language patterns. It utilizes speech-to-text conversion to analyze potential bullying cues within voice messages. YOLOv8, an object detection model, is employed to identify bullying hand gestures in uploaded images.

Data Collection and Storage: Gather text data from social media platforms, forums, messaging apps, etc., and images from various sources. Store the collected data securely in a database. Ensure compliance with privacy regulations.

Preprocessing: Cleanse and preprocess the text data by removing irrelevant information, stop words, and special characters. Resize and preprocess images to a suitable format for YOLOv8.

Text Analysis: Utilize natural language processing (NLP) techniques to analyze text data. Implement sentiment analysis and language models to detect aggressive, offensive, or bullying language. Use machine learning models trained on

International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 05 May 2024, pp: 840-845 www.ijaem.net ISSN: 2395-5252

labeled data to classify text into different categories like bullying, hate speech, etc.

Image Detection: Implement YOLOv8 algorithm for object detection in images. Fine-tune the YOLOv8 model to detect specific objects or patterns associated with cyberbullying or offensive content. Train the model using a dataset of annotated images containing examples of cyberbullying or inappropriate content.

Integration: Integrate the text analysis and image detection components into a unified system. Develop APIs or micro services for seamless communication between different modules.

IV.METHODOLOGY

By introducing each algorithm, we set the stage for their roles and strengths within the context of the mental health project, emphasizing their unique attributes and contributions. The project description mentions specific algorithms: Machine Learning for text analysis, YOLOv8 for hand gesture recognition, and HMM for voice analysis. Here's a general idea of how they might be trained.

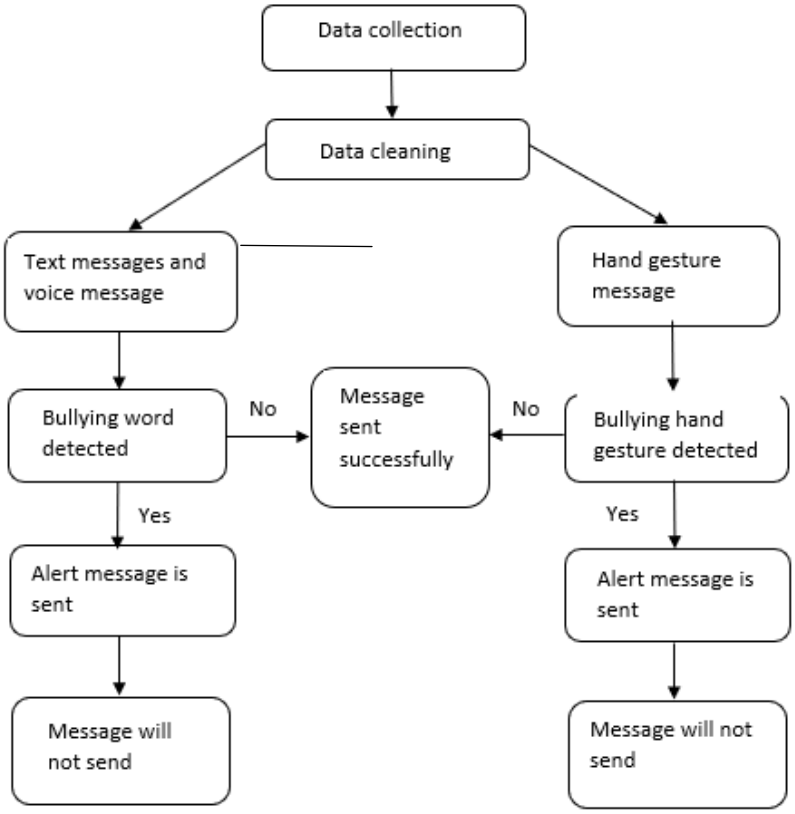

Fig. 2. Methodology

Step 1: Data Collection

Gather a comprehensive dataset from social media platforms, focusing on text, user meta data, and interactions indicative of cyberbullying behavior. Data should include a range of cyberbullying scenarios, capturing varying degrees of severity. Consider in corporating publicly available datasets and partnering with platforms to access real-world examples of cyberbullying incidents.

Step 2: Data Preprocessing

Clean and normalize the data to ensure consistency across different sources. This step involves handling missing or incomplete information, filtering out noise and irrelevant data, and standardizing text formatting. Additional processing may include removing offensive content, normalizing text, and balancing the dataset to avoid bias.

Step 3: Data Encoding

Use text-based encoding methods to convert raw text into numerical representations

International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 05 May 2024, pp: 840-845 www.ijaem.net ISSN: 2395-5252

that can be processed by machine learning algorithms. Techniques such as word embedding, one-hot encoding, or term frequency-inverse document frequency (TF-IDF) can be used to represent wordsorphrases. This stage transforms categorical information into a format suitable form achine learning.

Step 4: Model Training

Select and train appropriate machine learning algorithms, such as logistic regression, decision trees, or neural networks, to classify cyberbullying and assess its severity. Model tuning involves adjusting hyperparameters for optimal performance, focusing on metrics like accuracy, precision, recall, and F1-score.

Step 5: Model Testing

Split the dataset into training and testing sets to validate the model's effectiveness. Apply k-fold cross-validation for more robust assessment. Choose performance metrics such as accuracy and F1-score, and ensure the model performs well on unseen data. Test for bias and ensure the model's fairness across different groups and scenarios.

Step 6: Results

The trained model predicts cyberbullying incidents and assigns severity levels based on the input data. Predictions guide real-time monitoring and alerting systems, allowing moderators or administrators to take appropriate action. The system should integrate with applications to enable real-world use, providing feedback to users and supporting preventive measures against cyberbullying. Instances of cyberbullying in social media posts, this model is trained using the cleaned training data. It makes use of its capacity for sequential memory to record dependencies in the text.

YOLO-v8 Algorithm for image detection:

- **Gesture Datasets:** Gather datasets containing images or videos depicting various gestures associated with cyberbullying behaviors. These datasets may include examples of offensive gestures, threatening postures, and other nonverbal cues indicative of cyberbullying.
- **Annotation:** Annotate the collected data to label specific gestures and actions relevant to cyberbullying, ensuring that the dataset is appropriately labeled for training and evaluation purposes.
- **Image Processing:** Preprocess the collected images to standardize their size, resolution, and format, facilitating consistent input for the YOLO model.
- **Data Augmentation**: Augment the dataset by applying transformations such as rotation, scaling, and flipping to increase its diversity and robustness, enhancing the model's ability to generalize to unseen data.
- **YOLO Architecture:** Implement the YOLO object detection architecture, which enables real-time detection of objects and gestures in images or videos with high accuracy and efficiency.
- **Transfer Learning:** Fine-tune a pre-trained YOLO model using the annotated gesture dataset, leveraging transfer learning to adapt the model to the specific task of cyberbullying gesture detection.
- **Performance Metrics:** Evaluate the trained YOLO model using standard object detection metrics such as precision, recall, and mean average precision (mAP) to assess its accuracy and effectiveness in detecting cyberbullying gestures.
- **Cross-Validation:** Perform cross-validation experiments to validate the model's performance across different subsets of the dataset, ensuring its robustness and generalizability.

V.RESULTS AND DISCUSSION

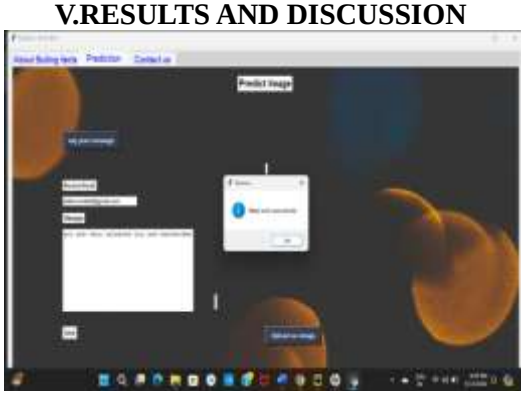

Fig 3. Email sent successfully.

The message sent by user to receiver is sent successfully. Because message does not contain any bullying words so message is sent to recipient’s email address.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 05 May 2024, pp: 840-845 www.ijaem.net ISSN: 2395-5252

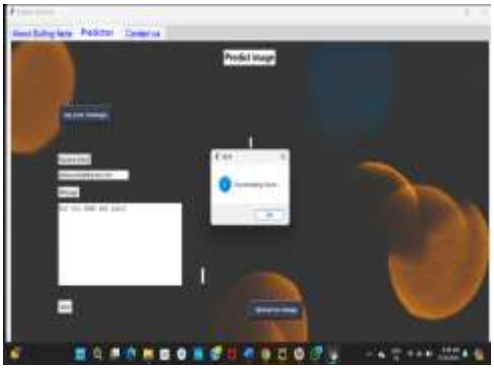

Fig 4 Bullying word detected.

User is sending some bullying words in message to the receiver. The system will analyses that bullying word and gives the alert and warning message to user that bullying word is found in message and message will not be sent to the receiver. If the user types a text without using bullying words then message will be send to receivers email.

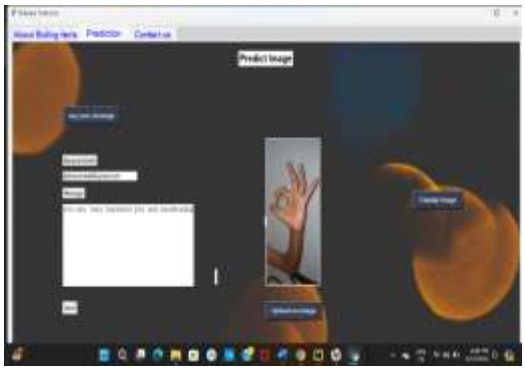

Fig 5 Upload any hand gesture image.

User can upload a hand gesture image as message to receiver. User selected any gesture from system send as a message to receiver and allow the model to classify image.

Fig 6 Classify the image with accuracy.

The uploaded hand gesture accuracy of sending image to receiver. If the hand gesture is without bulling sign then it will be sent to receiver. The snapshot indicating as ok with some accuracy because hand gesture does not contain bulling sign.

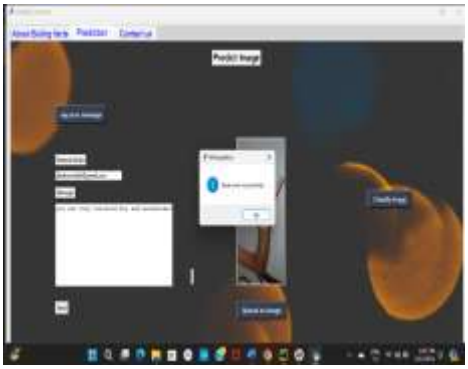

Fig 7 Email sent successfully.

The hand gesture image sent by user to receiver is sent successfully. Because message sent as hand gesture image does not contain bullying sign so message as image is sent to recipient's email address.

VI.CONCLUSION

Social media platforms are rife with both positive and negative interactions. While they offer opportunities for connection and expression, they can also be breeding grounds for cyberbullying.This project tackles cyberbullying on social media by employing a multifaceted approach. It analyzes text messages for bullying language using a Hidden Markov Model (HMM). If potential bullying words are detected, the sender receives an alert highlighting the issue.the project incorporates hand gesture recognition using YOLOv8 to identify bullying gestures in uploaded images. If a bullying gesture is identified, the image is blocked from being sent. This dual approach aims to address both verbal and non-verbal forms of cyberbullying within social media communication.This project contributes to a safer social media environment by empowering users to identify and potentially prevent their own bullying behavior before messages are sent.

VII FUTURE SCOPE

Develop the ability to differentiate between friendly teasing and bullying based on the relationship between the sender and receiver.Expand the system's capabilities to detectbullying language inmultiple languages.Consider cultural nuances of communication styles whendetecting bullying across different languages.Develop audio and video analysis capabilities to detect bullying

behavior in video content.Develop parental control features that allow parents to monitor their children's online activity and receive alerts about potential bullying situations.

VIII. ACKNOWLEDGMENT

We extend our heartfelt thanks to the Principal Dr. Yuvaraju B.N and the management and staff of PES Institute and Technology for providing us with necessary insights that enabled us to implement our project. Words cannot describe our gratitude to Dr. Arjun U , HoD, Department of CSE, PES Institute and Technology for his constant support throughout the journey.

REFERENCES

[1]. Mohammed Al-Hashedi, Lay-Ki Soon, Hui-NGO Goh, Amy Hui Lan Lim, and eu- Gene Siew: Cyberbullying Detection Based on Emotion (2023).

[2]. Mohammed Hussein Obaid, Shawkat Kamal Guirguis and Saleh Mesbah elkaffas: Cyberbullying Detection and Severity Determination Model (2023).

[3]. Teoh Hwai Teng and Kasturi Dewi Varathan: Cyberbullying Detection in Social Networks: A Comparison Between Machine Learning and Transfer Learning Approaches (2023).

[4]. Krishanu Maity, Sriparna Saha: Emoji, Sentiment and Emotion Aided Cyberbullying Detection in Hinglish (2022).

[5]. Aparna Sankaran Srinath, Hannah Johnson, Gaby G. Dagher, and Min Long:Bully Net: Unmasking Cyberbullies on Social Networks (2021).

[6]. Antonio Calvo-Morata, Dan Cristian Rotaru, Cristina Alonso-Fernandez, Manuel Freire-Moran,Ivan Martınez-Ortiz: Validation of a CyberbullyingSerious Game Using Game Analytics (2020).

[7]. Rui Zhao and Kezhi Mao:Cyberbullying Detection based on Semantic-EnhancedMarginalized Denoising Auto-Encoder (2017).

[8]. Semiu Salawu, Yulan He, and Joanna LumsdenApproaches to Automated Detection of Cyberbullying: A Survey (2017).

[9]. L. P. D. Bosque and S. E. Garza: Prediction of Aggressive Comments in Social Media: an Exploratory Study (2016).

[10]. Belal Abdullah Hezam Murshed, Jemal Abawajy Suresha Mallappa, Mufeed Ahmed Naji Saif, and Hasib Daowd Esmail Al-ArikiDEA-RNN: A Hybrid Deep Learning Approach for Cyberbullying Detection in Twitter Social Media Platform(2022).