CHAPTER I
INTRODUCTION

1.1 CYBERBULLY

Cyberbullying, the deliberate use of digital communication tools like social media, texting, and email, is employed by individuals to harass, threaten, or emotionally harm others. Unlike traditional forms of bullying, cyberbullying extends beyond physical spaces, allowing perpetrators to reach victims at any time and from any location. It can take various forms, including sending menacing messages, spreading false rumors, disclosing personal information, impersonating someone, and even cyberstalking [7]. This form of harassment can leave victims feeling isolated, helpless, and violated [3], as the online nature of cyberbullying makes it both public and difficult to escape. Given the severe emotional and psychological impact it can have, recognizing and preventing cyberbullying is an essential societal responsibility.

The repercussions of cyberbullying on victims can be profound, affecting mental well-being and leading to conditions such as anxiety [2], depression, and low self-esteem. In extreme cases, the emotional toll can drive individuals to self-harm or even suicide. As digital spaces become more integrated into everyday life, the potential for online harassment increases, highlighting the urgent need to address this issue effectively. Unlike physical bullying, which may have clear visual signs [1], cyberbullying is often hidden, making it difficult for parents, educators, and peers to detect and intervene. Thus, the stakes are high, as prolonged exposure to cyberbullying can have long-term consequences on a person's psychological health and quality of life.

Cyberbullying detection aims to identify instances of online harassment by monitoring digital communication channels and highlighting harmful interactions. Traditionally, this task has been handled by human moderators who review online content and identify potential instances of bullying [11,8]. However, the vast amount of data generated daily on social media, forums, and other online platforms makes it challenging for human moderators to keep up. Manual review alone is therefore insufficient for addressing the scale [3] of cyberbullying, prompting a shift toward automated detection systems that can process and analyze vast amounts of digital content in real-time [4]. Machine learning and artificial intelligence (AI) have become powerful tools for automating cyberbullying detection [6]. By training algorithms on large datasets containing labeled examples of cyberbullying and non-cyberbullying content, these systems can learn to identify patterns and characteristics commonly associated with harmful interactions.

Once trained, these models can analyze new posts, messages, and comments, flagging potential instances [8] of cyberbullying for further review. This approach enables rapid and large-scale monitoring of online content, helping to identify problematic behaviour more efficiently than manual methods alone.

Automated detection systems can take different actions upon identifying potential cyberbullying instances. In some cases, the system may flag content for review by a human moderator, who can verify the findings and decide on further action [2]. Alternatively, some platforms allow detection tools to automatically block or delete offensive content, thereby limiting the reach of harmful material [5]. While this can be effective in curbing cyberbullying on a larger scale, it also raises concerns about accuracy [1][3], fairness, and user privacy. It is essential for these systems to be highly accurate, as false positives could result in unfair censorship, while false negatives may allow harmful content to go undetected.

Despite their potential, automated cyberbullying detection systems face challenges in achieving consistent accuracy and fairness [17]. Language is nuanced, and understanding context is vital to distinguishing between harmful and benign interactions. Therefore, continuous improvements in algorithms and the use of contextual models like BERT can enhance the effectiveness of cyberbullying detection. However, developers [1] must also balance this with ethical considerations, ensuring that detection systems are unbiased and respect individual privacy. By refining these systems and addressing their limitations, automated cyberbullying detection can be a powerful tool in creating safer online spaces.

1.2 DEEP LEARNING

Deep learning is a subset of machine learning, which is essentially a neural network with three or more layers. These neural networks attempt to simulate the behaviour of the human brain—allowing it to ―learn‖ [2][4] from large amounts of data.

While a neural network with a single layer can still make approximate predictions, additional hidden layers can help to optimize and refine for accuracy. The Deep learning Architecture is shown in the fig 1.1

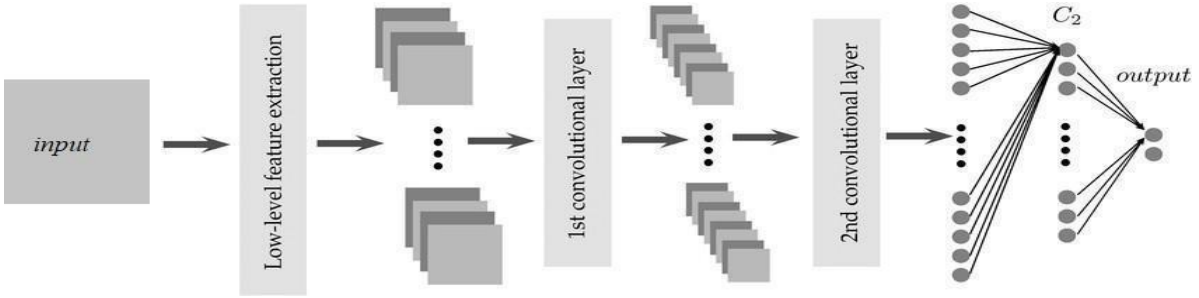


Figure 1.1 Deep learning Architecture

In deep learning, a computer model learns to perform classification tasks directly from images, text, or sound. Deep learning models can achieve state-of-the-art accuracy, sometimes exceeding human-level performance. Models are trained by using a large set of labeled data and neural network architectures that contain many layers. It is a field that is based on learning and improving on its own by examining computer algorithms [5]. However, advancements in Big Data analytics have permitted larger, sophisticated neural networks, allowing computers to observe, learn, and react to complex situations faster than humans. Deep learning has aided image classification, language translation, speech recognition. It can be used to solve any pattern recognition problem and without human intervention.

1.2.1 DEEP LEARNING & MACHINE LEARNING

Deep learning is a specialized form of machine learning. A machine learning workflow starts with relevant features being manually extracted from images. The features are then used to create a model that categorizes the objects in the image. With a deep learning workflow, relevant features are automatically extracted from images. In addition, deep learning performs ―end-to-endlearning‖ [11], where a network is given raw data and a task to perform, such as classification, and itlearns how to do this automatically. Another key difference is deep learning algorithms scale withdata, whereas shallow learning converges. A key advantage of deep learning networks is that theyoften continue to improve as the size of your data increases. The difference in working of machineand deep learning is shown in fig 1.2.

In machine learning, a programmer must intervene directly in the action for the model to come to a conclusion. In the case of a deep learning model, the feature extraction step is completely unnecessary. The model would recognize these unique characteristics of a car and make correct predictions. Deep Learning models [11] tend to increase their accuracy with the increasingamount of training data.

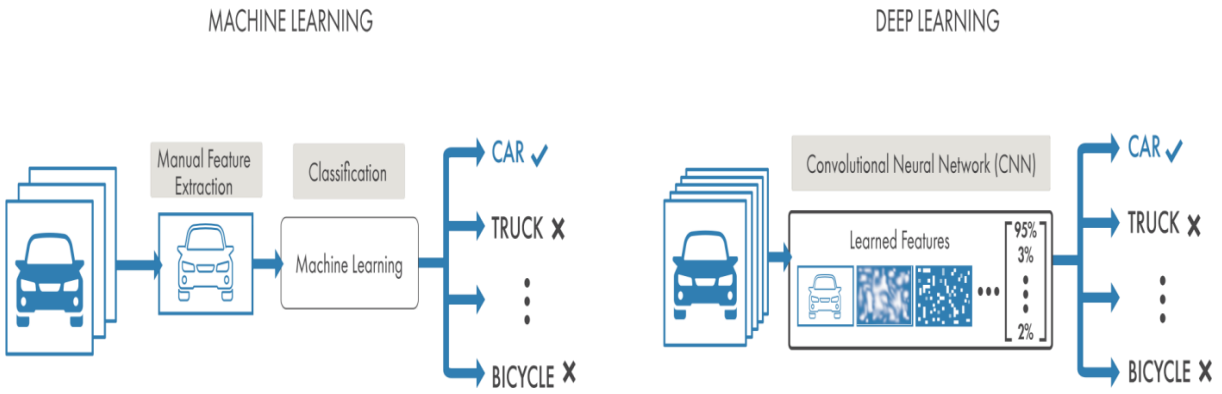


Figure 1.2 Machine Learning & Deep Learning

1.2.2 NEURAL NETWORKS

Deep learning algorithms attempt to draw similar conclusions as humans would by continually analyzing data with a given logical structure. To achieve this, deep learning uses a multi-layered structure of algorithms called neural networks [6]. The design of the neural network is based on the structure of the human brain[21]. Just as we use our brains to identify patterns and classify different types of information, neural networks can be taught to perform the same tasks on data. The Neural network[8] architecture is shown in fig1.3

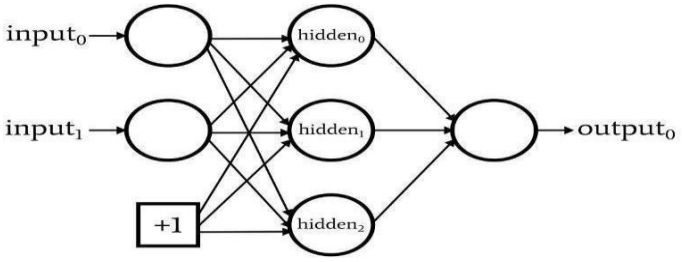


Figure 1.3 Neural Network Architecture

The individual layers of neural networks can also be thought of as a sort of filter that works fromgross to subtle, increasing the likelihood of detecting and outputting a correct result. The human brain works similarly. Whenever we receive new information, the brain tries to compare it with known objects [6]. The same concept is also used by deep neural networks. Neural networks enable us to perform many tasks, such as clustering, classification or regression[14]. With neural networks, we can group or sort unlabeled data according to similarities among the samples in this data.

1.2.3 TYPES OF DEEP NEURAL NETWORKS

- Multi-Layer Preceptrons (MLP)
- Convolutional Neural Networks (CNN)
- Recurrent Neural Networks (RNN)

1.2.4 MULTI-LAYER PERCEPTRONS(MLP)

A multilayer perceptron (MLP) is a class of a feedforward artificial neural network (ANN). MLPs models are the most basic deep neural network, which is composed of a series of fully connected layers. Today, MLP[3] machine learning methods can be used to overcome the requirement of high computing power required by modern deep learning architectures[3]. Each new layer is a set of nonlinear functions of a weighted sum of all outputs (fully connected) fromthe prior one.

1.2.5 CONVOLUTIONAL NEURAL NETWORK(CNN)

A convolutional neural network[7] (CNN, or ConvNet) is another class of deep neural networks. CNNs are most commonly employed in computer vision [4]. Given a series of images or videos from the real world, with the utilization of CNN, the AI system learns to automatically extract the features of theseinputs to complete a specific task, e.g., image classification , face authentication, and image semanticsegmentation[12]. Different from fully connected layers in MLPs, in CNN models, one or multiple convolution layers extract the simple features from input by executing convolution operations. Each layer is a set of nonlinear functions of weighted sums at different coordinates of spatially nearby subsets of outputs from the prior layer, which allows the weights to be reused.

AlexNet : For image classification, as the first CNN neural network to win the ImageNetChallenge in 2012, AlexNet consists of five convolution layers[17] and three fully connected layers. Thus, AlexNet requires 61 million weights and 724 million MACs (multiply-add computation) to classify the image with a size of 227×227.

VGG-16 : To achieve higher accuracy, VGG-16[13]is trained to a deeper structure of 16 layers consisting of 13 convolution layers and three fully connected layers, requiring 138 million weights and 15.5G MACs to classify the image with a size of 224×224.

GoogleNet : To improve accuracy while reducing the computation of DNN inference, GoogleNet introduces an inception module composed of different sized filters. As a result, GoogleNet achieves a better accuracy performance than VGG-16[13] while only requiring seven million weights and 1.43G MACs to process the image with the same size.

ResNet : The state-of-the-art effort, uses the ―shortcut‖ structure to reach a human-level accuracy with a top-5 error rate below 5%. In addition, the ―shortcut‖ module is used to solve the gradient vanishing problem during the training process, making it possible to train a DNN model[16] with a deeper structure.

1.2.6 ADVANTAGES OF CNN

CNN learns the filters automatically without mentioning it explicitly. These filters help in extracting the right and relevant features from the input data Advantages of Convolution Neural Network [4].

CNN captures the spatial features from an image. Spatial features[16] refer to the arrangement of pixels and the relationship between them in an image. They help us in identifying the object accurately, the location of an object, as well as its relation with other objects in an image[4].

1.2.7 RECURRENT NEURAL NETWORK (RNN)

A recurrent neural network (RNN) is another class of artificial neural networks [5] that use sequential data feeding. RNNs have been developed to address the time-series problem of sequential input data[5]. The input of RNN consists of the current input and the previous samples. Therefore, the connectionsbetween nodes form a directed graph along a temporal sequence[5].

1.2.8 ADVANTAGES OF RNN

RNN captures the sequential information present in the input data i.e. dependency between the words in the text while making predictions[5].

RNNs share the parameters across different time steps. This is popularly known as Parameter Sharing. This results in fewer parameters to train and decreases the computational cost.

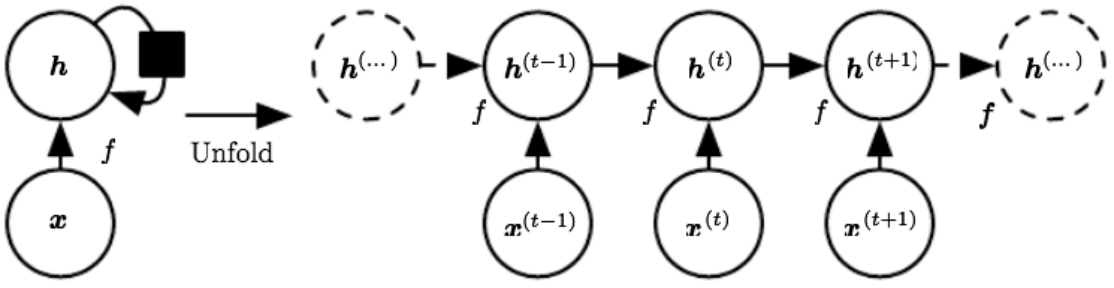


Figure 1.4 Unrolled RNN

1.2.9 BERT

Bidirectional Encoder Representations from Transformers [3], commonly known as BERT, is a groundbreaking model in natural language processing (NLP) developed by Google. Introduced in 2018, BERT marked a major shift in how language models understand context and meaning in text. Traditional NLP[3][7] models often processed language sequentially, either left-to-right or right-to-left, which limited their ability to capture the full context of a word based on its surrounding words.

However, BERT[3] uses a bidirectional approach, meaning it processes language in both directions simultaneously. This allows BERT to understand each word in the context of the words [3] that come both before and after it, greatly enhancing its ability to capture nuanced meanings and complex sentence structures.

The architecture of BERT is based on transformers, which are deep learning models that excel at processing sequential data by using attention mechanisms[11]. A transformer's attention mechanism enables it to weigh the importance of each word in a sentence relative to all other words, making it particularly effective for understanding long-range dependencies and contextual relationships[7].

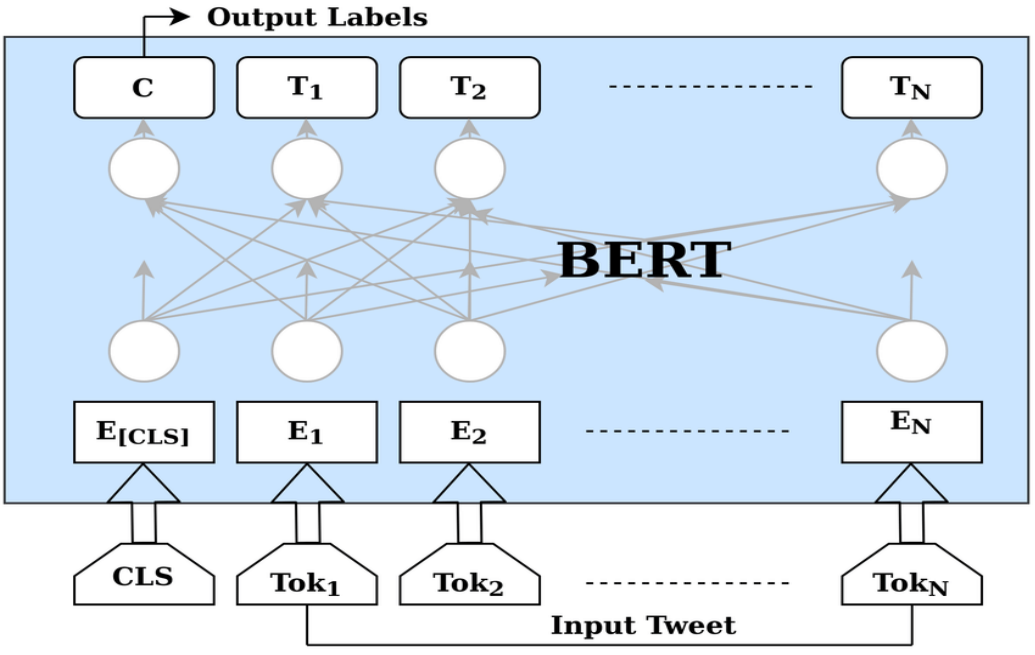


Figure 1.5 Architecture of BERT

In BERT's case, the model's architecture is pre-trained on a massive corpus of text data, such as Wikipedia and the BookCorpus [3][7], where it learns language patterns, word associations, and grammatical structures.

This pre-training phase allows BERT to develop a general understanding of language that can later be fine-tuned for specific tasks with relatively small amounts of task-specific data. One of the key innovations of BERT[3] is its use of ―masked language modeling‖ during pre-training. In this approach, some words in a sentence are randomly masked (hidden) from the model, and BERT is tasked with predicting the masked words based on their surrounding context. This forces the model to learn bidirectional context, as it must consider both preceding and succeeding words to make accurate predictions [7]. Additionally, BERT employs a technique called ―next sentence prediction,‖ where it learns the relationships between sentences by predicting if two sentences in a text are consecutive.

This combination of techniques enables BERT to excel at a variety of NLP tasks that require deep contextual understanding. After pre-training [3], BERT can be fine-tuned for a wide range of NLP applications, such as sentiment analysis, question answering, and named entity recognition. Fine-tuning involves training BERT[4] on a smaller, labeled dataset related to the specific task, allowing it to adapt its general language understanding to the particular nuances of the task at hand.

Because of its rich pre-trained language knowledge, BERT often requires minimal fine-tuning to achieve state-of-the-art performance on many NLP[7] benchmarks. This versatility makes BERT particularly valuable in the NLP field, as it can be adapted to new applications with relative ease and high accuracy.

BERT's impact on NLP has been profound, raising the bar for what language models can achieve and inspiring a new wave of research in transformer-based models. The success of BERT led to the development of other transformer-based models, like GPT [11], RoBERTa, and T5, each with unique optimizations and applications.

BERT's bidirectional approach has set a standard for understanding context in language, and its architecture has proven highly adaptable, influencing advancements in text generation, translation, and summarization. Moreover, BERT has opened up possibilities for handling complex NLP challenges that require contextual sensitivity [7][11], such as detecting sarcasm, identifying sentiment, and understanding conversational nuances.

CHAPTER II
LITERATURE SURVEY

2.1 LITERATURE SURVEY - I

Title: Cyberbullying Detection in Social Networks: A Comparison Between Machine Learning and Transfer Learning Approaches.

Author : Ainize Martínez Soto, Cristina Lopez-del Burgo, Aranzazu Albertos.

Year: 2024

Description:

Their research developed an automatic system for detecting cyberbullying, using two approaches: Conventional Machine Learning (CML) and Transfer Learning. The CML method analyzed features like textual content, sentiment, emotional indicators, and toxicity levels using a Logistic Regression model, achieving an F-measure of 64.8%. Their study highlighted the importance of incorporating psycholinguistic tools like LIWC 2022 and Empath's lexicon, which provide deeper insights into the emotional and psychological aspects of language used in cyberbullying. By combining these advanced features with traditional machine learning techniques, the system improved detection performance. Their findings suggest that integrating linguistic and psychological characteristics enhances the accuracy of cyberbullying detection, contributing to safer online environments[3].

Conclusion:

We conclude that the effectiveness of combining traditional machine learning will improve the cyberbullying detection by integrating features like sentiment, emotional indicators, and toxicity with tools such as LIWC 2022 and Empath's lexicon, and this method will enhance the accuracy. The findings emphasize the value of considering both linguistic and psychological factors, contributing to more reliable tools for early cyberbullying detection and fostering safer online environments.

2.2 LITERATURE SURVEY - II

Title: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.

Author: Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova.

Year: 2022

Description:

Their Research work explores the application of BERT (Bidirectional Encoder Representations from Transformers) in natural language processing tasks. BERT has revolutionized the NLP field by introducing a pre-trained transformer-based model that captures deep bidirectional context, allowing for significant improvements in various downstream tasks. Their study delves into the architecture of BERT, detailing its transformer mechanism and the use of masked language modeling and next sentence prediction for pre-training. Their research evaluates the effectiveness of fine-tuning BERT for specific tasks, demonstrating enhanced performance in tasks such as sentiment analysis, question answering, and text classification. Their work concludes with a discussion on the limitations and potential areas for improvement in BERT-based models, including challenges in computational efficiency and the handling of out-of-distribution data[22].

Conclusion:

We conclude that the transformer-based models for understanding language context will improve the efficiency of cyberbully detection. BERT's bidirectional architecture and pre-training techniques enable superior performance across various tasks. Despite its success, challenges remain, particularly regarding computational demands.

2.3 LITERATURE SURVEY - III

Title: NLP techniques for automating responses to customer queries: a systematic review.

Author: Peter Adebowale Olujimi, Abejide Ade-Ibijola.

Year: 2021

Description:

The demand for automated customer support has surged in recent years, driven by advancements in Natural Language Processing (NLP). Conversational AI now enables chatbots to understand and respond to customer inquiries without human intervention, improving efficiency across sectors like banking, healthcare, education, and manufacturing. Their study systematically reviewed 73 articles from reputable sources, analyzing the application of NLP techniques in automating customer service. The findings provide a comprehensive overview of prior research, highlighting benefits, existing practices, and future research directions. Key implications and recommendations for business applications of NLP are also discussed[23].

Conclusion:

We conclude that integrating an automated response system will substantially enhance the effectiveness of BERT's automated detection system. Their enhancement stems from the system's ability to provide instant notifications and streamline the communication of detection results. By ensuring timely responses and reducing manual intervention, the system will improve operational efficiency and reliability. Additionally, the automated response mechanism can facilitate continuous monitoring and quicker decision-making.

2.4 LITERATURE SURVEY - IV

Title: Detecting cyberbullying using deep learning techniques: a pre-trained glove and focal loss technique.

Author: Amr Mohamed El Koshiry, Entesar Hamed, I.Eliwa, Tarek Abd El-Hafeez.

Year: 2024

Description:

Their Research compares the performance of classical and deep learning algorithms in detecting cyberbullying, using metrics like accuracy, precision, recall, and F1 score. The Focal Loss algorithm achieved the highest accuracy and precision, but most algorithms showed low recall, indicating many cyberbullying instances went undetected. To address this, their study proposes a hybrid model combining Convolutional Neural Networks (CNN) and Bidirectional Long Short-Term Memory (Bi-LSTM) layers. The model extracts spatial features with CNN and captures temporal dependencies with Bi-LSTM. Trained on pre-processed tweet data with GloVe embeddings and using focal loss to tackle class imbalance, the model aims to improve both accuracy and recall, enhancing cyberbullying detection across diverse data inputs[21].

Conclusion:

We concluded to implement the Focal Loss algorithm due to its high accuracy in handling class imbalance issues and propose a hybrid model combining CNN and Bi-LSTM layers for enhanced performance. The Focal Loss algorithm excels in focusing on hard-to-classify samples, making it a valuable addition to the model's architecture. By integrating CNN layers for feature extraction and Bi-LSTM layers for capturing sequential dependencies, the hybrid model leverages the strengths of both techniques. Their approach aims to achieve robust learning, improved generalization, and superior performance across diverse datasets. The combination is particularly promising for tasks requiring both spatial and temporal feature analysis.

2.5 LITERATURE SURVEY-V

Title: Social Sensor for Real Time Event Detection.

Author: Prajakta Patil, Pooja Patil, Amruta Salvi, Prof. Mrunmayee Hatiskar.

Year: 2021

Description:

Social networking sites, especially Twitter, are vital for sharing and tracking real-time events like earthquakes and traffic. Our system analyzes tweets using a classifier to detect events based on keywords, word count, and location. Users act as sensors, and a priority-based algorithm identifies events and their approximate locations. For traffic management, factors like volume, speed, and road occupancy are analyzed. Tweets are categorized, and a probabilistic model aids event detection. Grouping users by interests, such as traffic or weather, enhances event tracking and response through real-time data processing[16].

Conclusion:

We conclude that our current system in effectively classifying data by introducing a Priority-based Message Stack Queuing Algorithm. Their approach aims to efficiently classify positive and negative tweets, enabling the development of a probabilistic model for accurate event detection. By prioritizing messages based on relevance and sentiment, the algorithm ensures timely processing of critical information. The use of this method not only improves classification accuracy but also enhances the system's responsiveness to real-time events.

2.6 LITERATURE SURVEY COMPARISON

S.NO	TITLE	AUTHOR	YEAR	TECHNIQUE USED	CONCLUSION
01.	Cyberbullying Detection in Social Networks: A Comparison Between Machine Learning and Transfer Learning Approaches.	Ainize Martínez Soto, Cristina Lopez-del Burgo, Aranzazu Albertos.	2024	Conventional Machine Learning (CML), Logistic Regression	We concluded that, by Combining traditional machine learning with sentiment, emotional indicators, and tools like LIWC, will enhances the cyberbullying detection accuracy.
02.	BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.	Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova.	2022	BERT, NLP	We concluded that the effective transformer-based models like BERT for better understanding language context, and for its superior performance due to their bidirectional architecture and pre-training.
03.	NLP techniques for automating responses to customer queries: a systematic review.	Peter Adebowale Olujimi, Abejide Ade-Ibijola.	2021	RNN, RoBERT, Keyword Extraction, NMT	We concluded, that the Automated responses system will en-large the effectiveness of the Automated Detection System of BERT.

04.	Detecting cyberbullying using deep learning techniques: a pre-trained glove and focal loss technique.	Amr Mohamed El Koshiry, Entesar Hamed, I.Eliwa, Tarek Abd El-Hafeez.	2024	Bi-LSTM, CNN	We concluded to implement the Focal Loss algorithm's high accuracy and proposes a hybrid model combining CNN and Bi-LSTM layers for better performance.
05.	Social Sensor for Real Time Event Detection.	Prajakta Patil, Pooja Patil, Amruta Salvi, Prof. Mrunmayee Hatiskar.	2021	Priority-based Message Stack, Particle Filtering	We concluded, that the real time detection of bully comments, by using the Priority-based message stack queuing algorithm to classifies positive and negative tweets to develop a probabilistic model for event detection.

Table 2.6 Review on recent research on cyberbullying detection

CHAPTER III
SYSTEM STUDY

3.1 EXISTING SYSTEM

Existing systems for cyberbullying detection, particularly in social networking environments like Facebook, face significant challenges [13] in identifying and mitigating online abuse. These platforms provide users with numerous avenues for positive interaction and communication; however, they can also foster an environment ripe for harmful behaviors such as cyberstalking [2], where individuals are targeted through ridicule, torment, and slander without any face-to-face interaction.

Traditional approaches often rely on manual moderation, which can be ineffective due to the sheer volume of content generated daily and the nuanced language used in abusive posts. As the prevalence of social media grows [14], there is an urgent need for automated systems capable of detecting and removing harmful content to prevent the potential for widespread psychological harm among victims [17]. To address these challenges, recent research has focused on developing advanced machine learning (ML) models specifically tailored for cyberbullying detection in languages such as Bengali [2].

The proposed hybrid ML model, known as Bengalibullying, emphasizes effective text preprocessing techniques to transform raw Bengali text data into a usable format, facilitating better analysis and classification. By employing the TfidfVectorizer [8] for feature extraction, the model captures essential information from the text while ensuring that it can differentiate between benign and abusive content. Furthermore, the Instance Hardness Threshold (IHT) resampling technique [8] is implemented to balance the dataset, thus avoiding overfitting or underfitting, which are common pitfalls in machine learning models.

The use of a large publicly available Bangla text dataset consisting of 44,001 comments demonstrates the model's ability to achieve impressive performance metrics, with accuracy rates reaching 68.57% [2] in binary classification and 68.82% in multilabel classification. This surpasses previous efforts in the field, showcasing the model's effectiveness in accurately identifying and categorizing instances of cyberbullying in Bengali [2], ultimately contributing to a safer online environment for users.

3.2 EXISTING SYSTEM ARCHITECTURE

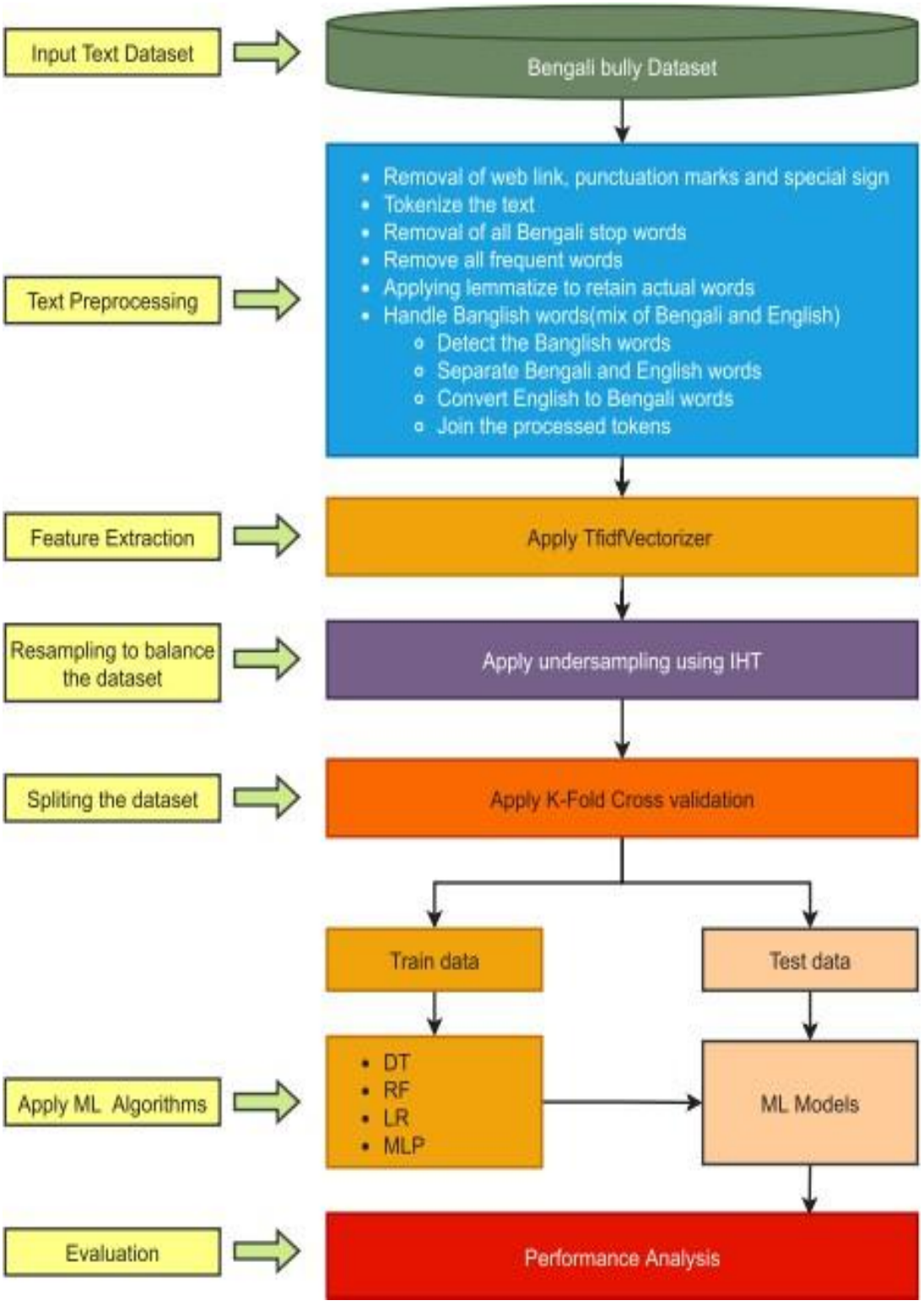


Figure 3.1 Bengalibullying detection architecture.

The architecture diagram for the hybrid machine learning model designed for cyberbullying detection outlines a structured approach to identifying abusive content in Bengali on social media platforms[16][17]. It begins with data collection, where raw Bengali text data is gathered from various social media sources. This data undergoes preprocessing steps, including cleaning and normalization, to ensure it is in a usable format.

Following preprocessing, feature extraction is performed using the TfidfVectorizer, which transforms the text into a structured representation that highlights the significance of words and phrases relevant to cyberbullying[3]. The extracted features [18] are then utilized to train the hybrid model, which combines multiple machine learning algorithms to enhance classification accuracy.

Finally, the model outputs predictions, categorizing the input text as either bullying or non-bullying, thereby enabling automated detection and intervention. This architecture not only improves the reliability of cyberbullying[3] detection but also addresses the linguistic nuances [9] of the Bengali language, contributing to more effective online safety measures.

3.3 LIMITATIONS IN EXISTING SYSTEM

1. **Generalizability Issues:** The focus is primarily on the Bengali language, which limits the applicability of the methodology to other languages and cultural contexts.
2. **Dataset Constraints:** The study uses a publicly accessible dataset of 44,001 comments, which may not represent the full range of cyberbullying scenarios or the diversity of online interactions.
3. **Limited Exploration of Advanced Models:** The study primarily uses traditional machine learning models like Decision Trees, Random Forests, Logistic Regression, and Multilayer Perceptrons. It does not explore advanced deep learning models or transformers such as BERT or RoBERTa, which could potentially improve detection accuracy.
4. **Class Imbalance:** While the Instance Hardness Threshold (IHT) technique is used to address dataset imbalance, it reduces the data volume for certain classes, which might affect the comprehensiveness of the results.
5. **Real-Time Implementation:** The study does not address the practical challenges of implementing this system in real-world scenarios, such as data privacy concerns, scalability, real-time processing, and handling evolving forms of cyberbullying.

CHAPTER IV
PROPOSED SYSTEM

4.1 PROPOSED SYSTEM INTRODUCTION

The proposed system utilizes the Bidirectional Encoder Representations from Transformers (BERT) model to improve cyberbullying detection through an in-depth understanding of language subtleties. BERT's bidirectional architecture allows the model to capture the entire context of sentences by analyzing words both preceding and following a target term.

This capability is crucial in distinguishing between benign and harmful communications, particularly in the context of cyberbullying, where the intent behind messages can be highly nuanced. By focusing on the broader context in which words are used, the system enhances its accuracy in identifying harmful interactions that might otherwise go unnoticed.

To further augment its effectiveness, the system is trained on a diverse dataset that includes texts in both Tamil and English, catering to the linguistic and cultural variations found in online communication.

This multilingual aspect ensures that the model is adaptable and relevant across different communities, significantly bolstering its capacity to detect various forms of cyberbullying. By integrating BERT’s sophisticated language processing strengths, the proposed system not only aims for high accuracy in predictions but also supports proactive measures for early intervention. Ultimately, this approach aspires to create a safer and more supportive online environment, mitigating the impact of cyberbullying on users.

4.2 PROPOSED SYSTEM ARCHITECTURE

` The architecture of the proposed cyberbullying detection system leverages the Bidirectional Encoder Representations from Transformers (BERT) model, designed to enhance understanding of language intricacies in both Tamil and English. The system starts with an input layer that collects raw text data from various social media platforms, comprising both bullying and non-bullying examples to create a balanced dataset.

This data then undergoes a pre-processing module, where noise such as special characters and stop words is removed, and techniques like tokenization and lemmatization are applied to ensure a standardized text format.

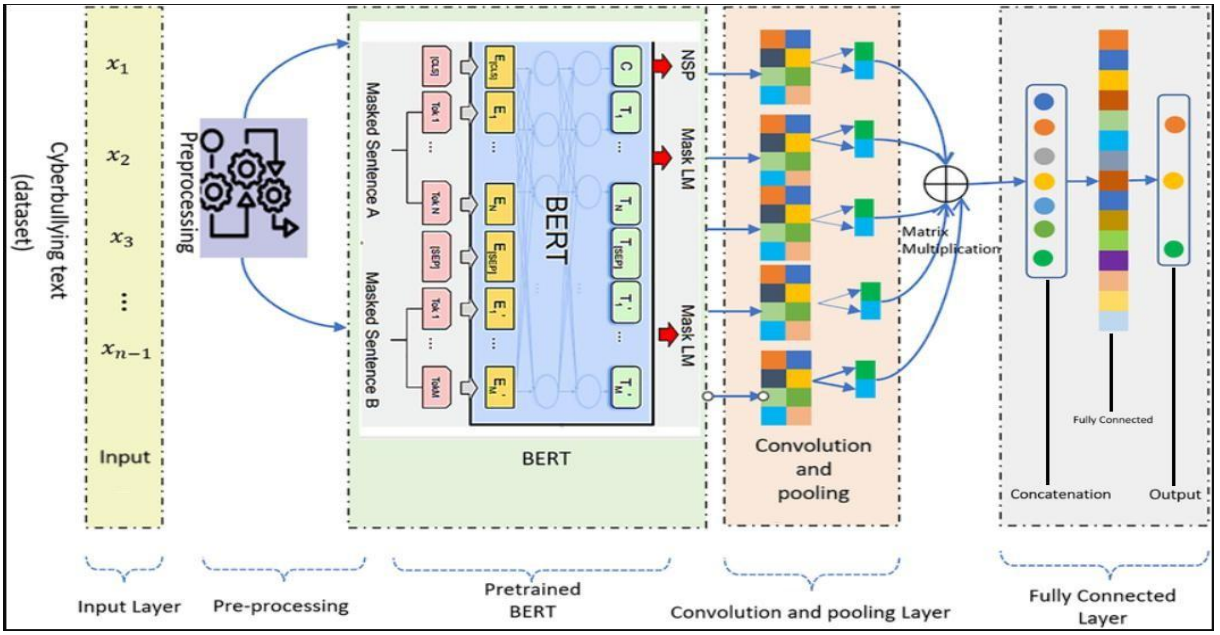


Figure 4.1 Proposed System architecture.

Once the text is cleaned, it is processed by the BERT model in the feature extraction phase. BERT's bidirectional processing capability allows it to analyze the context of words within sentences, capturing subtle nuances that are critical in distinguishing between benign and harmful messages. The output consists of rich contextual embeddings that reflect the meaning and intent behind the text. These features are subsequently fed into a classification layer, which employs machine learning algorithms to determine the likelihood of the text being classified as bullying.

Finally, the output layer presents the model's predictions, enabling real-time identification of harmful content and fostering a safer online environment. By integrating these components, the proposed architecture effectively addresses the complexities of cyberbullying detection, contributing to early intervention and support in online interactions.

4.3 ADVANTAGE OF PROPOSED SYSTEM

The primary advantage of this system is its utilization of BERT's sophisticated contextual understanding to detect bullying language with high accuracy, even across multiple languages such as Tamil and English. By analyzing text in a bidirectional manner, BERT captures the nuanced meanings and subtle intentions within sentences, which is critical for distinguishing between benign and harmful language. This capability is particularly valuable in multilingual settings, where language diversity and cultural nuances often pose significant challenges for traditional detection methods.

By incorporating linguistic variations, the system adapts effectively to the diverse nature of online communities, providing a precise and inclusive approach to identifying cyberbullying. Ultimately, this robust multilingual detection mechanism helps create a safer digital environment by reliably flagging harmful content and supporting timely intervention, contributing to the well-being of users across different linguistic backgrounds.

CHAPTER V

SYSTEM REQUIREMENTS

5.1 HARDWARE REQUIREMENTS

- Processor : Processor : Intel i7 upto 4.0 GHzSpeed
- RAM : 8 GB
- Hard Disk : 512 GB SSD

5.2 SOFTWARE REQUIREMENTS

- Language : Python - 3.7 or above
- Anaconda 2023.9 (Managing Python Packages)
- Database: Mysql 8.0

5.3 ENVIRONMENTAL SETUP SCREENSHOT

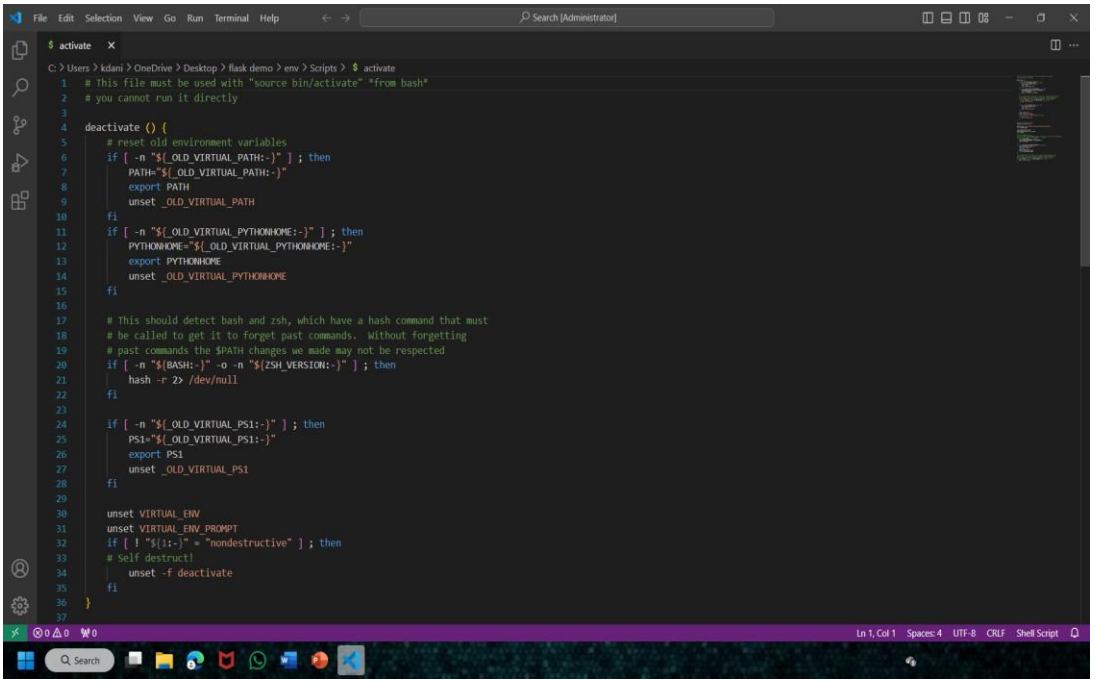


Figure 5.1 Environmental setup screenshot.

CHAPTER VI

DESIGN AND IMPLEMENTATION

6.1 PROPOSED SYSTEM ARCHICTURE

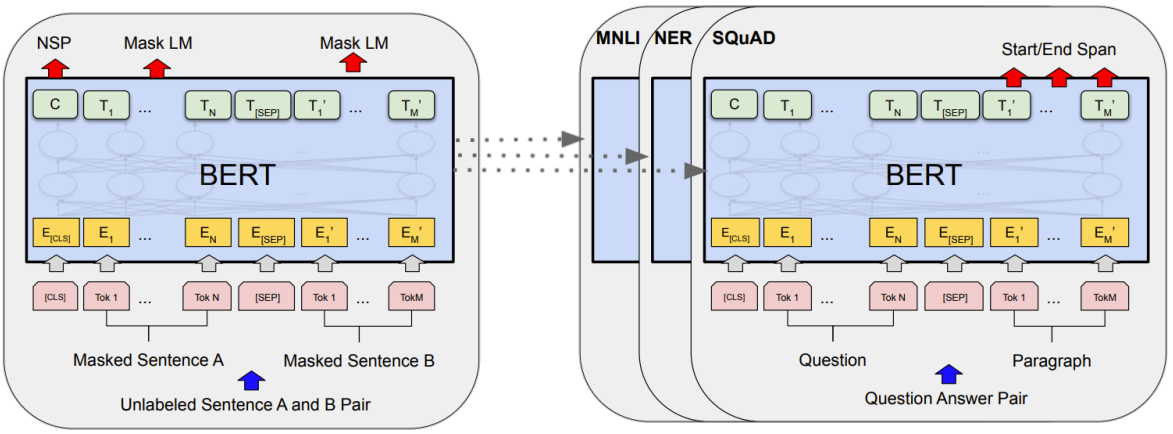


Figure 6.1 Pre-Training & Fine-Tuning of BERT Model.

6.2 IMPLEMENTATION

The Proposed system uses the BERT model to detect cyberbullying in Tamil and English text. It includes text preprocessing, feature extraction with BERT, and classification using machine learning. The model accurately identifies harmful content and supports real-time detection for safer online interactions.

Data Acquisition

Collecting raw textual data from various social media platforms, primarily focusing on comments and posts written in Bengali. This data includes both cyberbullying and non-cyberbullying content, which is labeled accordingly.

Data Preprocessing

The collected data undergoes a comprehensive preprocessing phase. This involves cleaning the text by removing noise such as special characters, emojis, URLs, and converting the text to lowercase. Tokenization is applied to split sentences into individual words or subwords, and stopwords common words that do not carry much semantic weight are removed. Additionally, lemmatization is used to reduce words to their base or root form, standardizing the text and preparing it for feature extraction.

Feature Extraction

The TF-IDF technique, which assigns weights to words based on how frequently they appear across different documents. This helps to highlight important terms while downplaying common ones. BERT's bidirectional attention mechanism analyzes the context of a word based on both its preceding and following words, generating deep semantic representations of the text that are far more informative than basic frequency-based methods.

Model Construction

Naïve Bayes is a fast and simple model based on probability and assumes feature independence. It works well for basic text classification but struggles with complex language patterns and context.

1. **SVM** finds the best boundary between classes using a hyperplane. It handles high-dimensional text data well but can be slower and less flexible for large, multilingual datasets.
2. **Logistic Regression** is a linear model that predicts class probabilities. It performs decently on structured features like TF-IDF but lacks depth for understanding nuanced text.
3. **Random Forest** uses multiple decision trees to improve accuracy and reduce overfitting. It handles noise and imbalance better than some models but still lacks context awareness.
4. **Multilingual BERT** is a deep learning model that understands word meanings using full sentence context in multiple languages. It gives the highest accuracy for detecting subtle and complex cyberbullying patterns.

Model Training

Multiple machine learning models are explored to find the most effective solution for cyberbullying detection. Traditional models like Logistic Regression, Random Forests, Naïve Bayes, and Support Vector Machines are trained using the TF-IDF features.

However, to improve accuracy and handle the complex linguistic nuances in Bengali, the system ultimately adopts a multilingual BERT model . This transformer-based architecture is fine-tuned on the dataset to optimize its performance for the cyberbullying detection task. The data is split into training and test sets to evaluate the model's ability to generalize, and performance is measured using metrics like accuracy and precision.

Evaluation

The trained model is tested on unseen data to measure its performance. Key metrics like accuracy, precision, recall, and F1 score are calculated to understand how well the model can detect cyberbullying. A good model should accurately classify both harmful and non-harmful texts.

6.3 MODULES USED IN PROPOSED SYSTEM

Data Collection Module

Responsible for gathering raw textual data from various social media platforms(e.g : facebook, youtube). This data includes examples of both bullying and non-bullying content to ensure the dataset is balanced and representative for training the detection system.

Data Preprocessing Module

The pre-processing module cleans and prepares the collected text data for further analysis. It removes noise such as special characters and stop words, and applies techniques like tokenization and lemmatization to convert the text into a standardized format suitable for feature extraction.

Feature Extraction Module

BERT model extract’s contextual embeddings from the pre-processed text. BERT’s bidirectional architecture allows it to consider the context of each word from both directions, making it highly effective in capturing subtle meanings and detecting nuanced cyberbullying language.

Classification Module

In this module, the features generated by BERT are used to classify the input text. Machine learning algorithms are applied to determine whether the content falls into the bullying or non-bullying category, based on patterns learned during training.

Automated Response System

Automatically generate alerts or responses based on detection results, enhancing the system's ability to intervene promptly and reduce manual monitoring efforts.

CHAPTER-VII
RESULT AND DISCUSSION

7.1 RESULT

The results confirm that the performance of various machine learning and deep learning models for cyberbullying detection in Bengali text. The study tested traditional models such as Multinomial Naive Bayes, Support Vector Machines, and Logistic Regression, alongside more advanced architectures like BiLSTM, multilingual BERT (mBERT) demonstrated the highest performance, achieving an accuracy of 95%, with similarly strong precision, recall, and F1-score metrics, all hovering around 94–95%.

Multilingual BERT also performed well, though slightly below mBERT, which is understandable given that BiLSTM is designed for broader multilingual coverage and not fine-tuned specifically for NLP. Traditional machine learning models, while computationally efficient, lagged significantly in terms of accuracy and contextual understanding, underlining their limitations in handling nuanced and context-dependent language patterns present in cyberbullying content.

7.2 DISCUSSION

The success of mBERT stems from its language-specific pretraining, which enables a deeper grasp of the linguistic subtleties in Multi-language. This highlights the critical role of using domain-specific or language-specific models when tackling NLP tasks in Multiple languages

1. The findings highlight the importance of using language-specific pre-trained models for tasks like cyberbullying detection in low-resource languages.
2. mBERT 's superior performance is attributed to its training on Bengali language data, capturing linguistic nuances better than BiLSTM or traditional models.
3. The project suggests that transformer-based models are more effective for this task than conventional machine learning algorithms.

7.3 ACCURACY METRICS

Model	Accuracy
Multinomial Naive Bayes	78%
Support Vector Machine	83%
Logistic Regression	82%
BiLSTM	89%
Multilingual BERT (mBERT)	98%
Bangla-BERT	95%

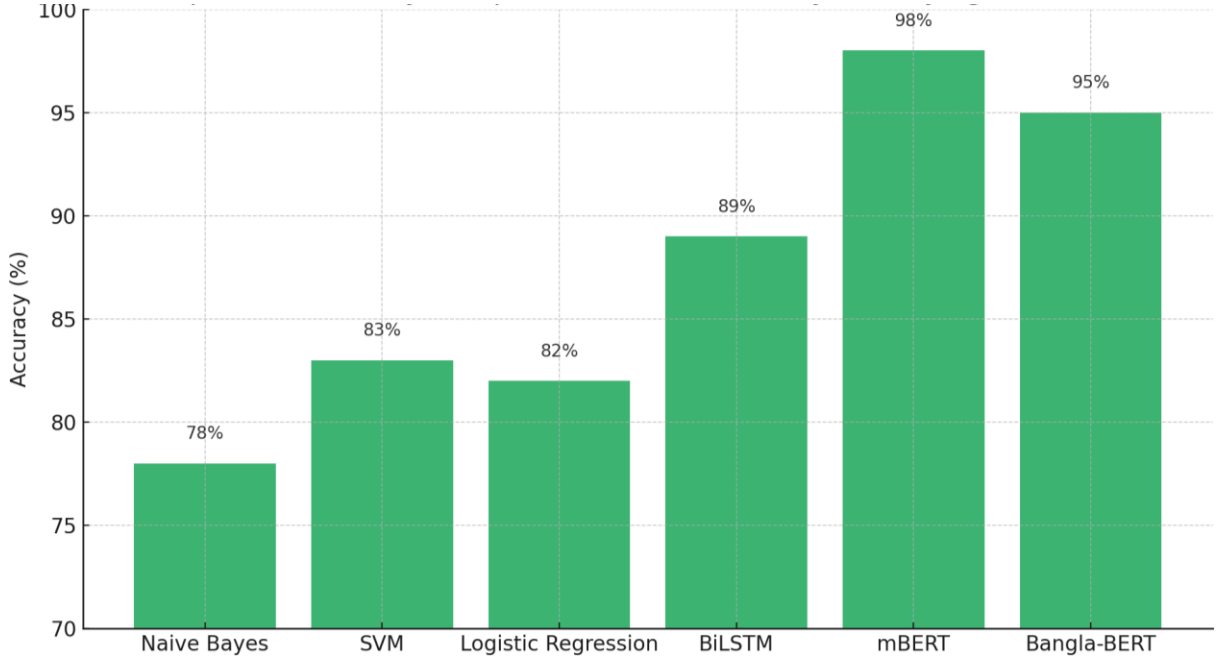


Figure 7.1 Accuracy metrics Graph

mBERT achieved the highest accuracy among all models, demonstrating its effectiveness in detecting cyberbullying in text due to its language-specific training. In contrast, traditional models like Naive Bayes and Logistic Regression showed comparatively lower accuracy, reflecting their limitations in capturing deep contextual information.

7.4 COMPARISON GRAPH :

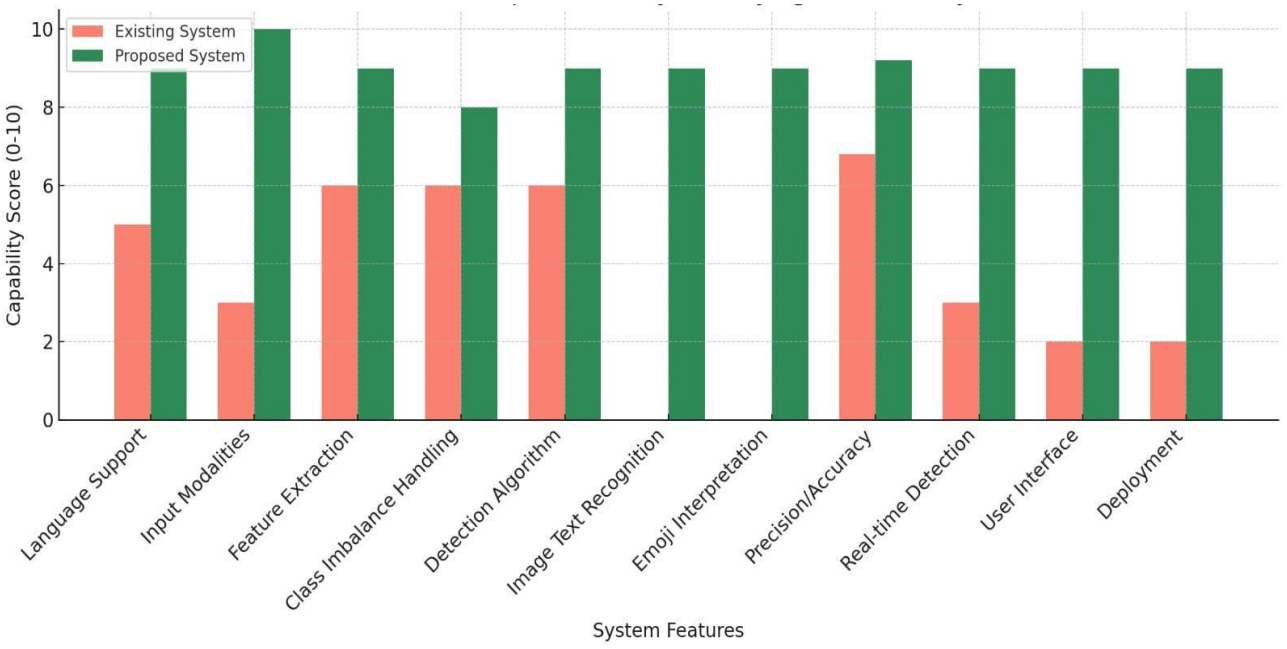


Figure 7.2 Comparison Graph of Existing & Proposed System

CHAPTER -VIII

CONCLUSION AND FUTURE ENHANCEMENT

8.1 CONCLUSION

This project successfully demonstrates the potential of a hybrid machine learning model in detecting bullying comments of social media comments with high accuracy,The main goal of this project was to detect cyberbullying in social media comments using advanced machine learning techniques especially the BERT model, which understands the meaning of words in context, performs much better than traditional machine learning models. It reads text in both directions (left-to-right and right-to-left), which helps it understand the true meaning of a sentence. This is very useful in identifying cyberbullying, where harmful messages can be subtle and depend on the tone or hidden context.

BERT was trained using a dataset containing both bullying and non-bullying comments in English. After training, the model could accurately detect abusive language, even when it was disguised or indirect. The results showed higher accuracy compared to earlier models like Naïve Bayes, SVM, and Logistic Regression, especially because those older models often miss context or sarcasm.

8.2 FUTURE ENHANCEMENT

Multilingual Expansion: Extend the model to support additional regional languages such as Tamil to cater to a wider user base.

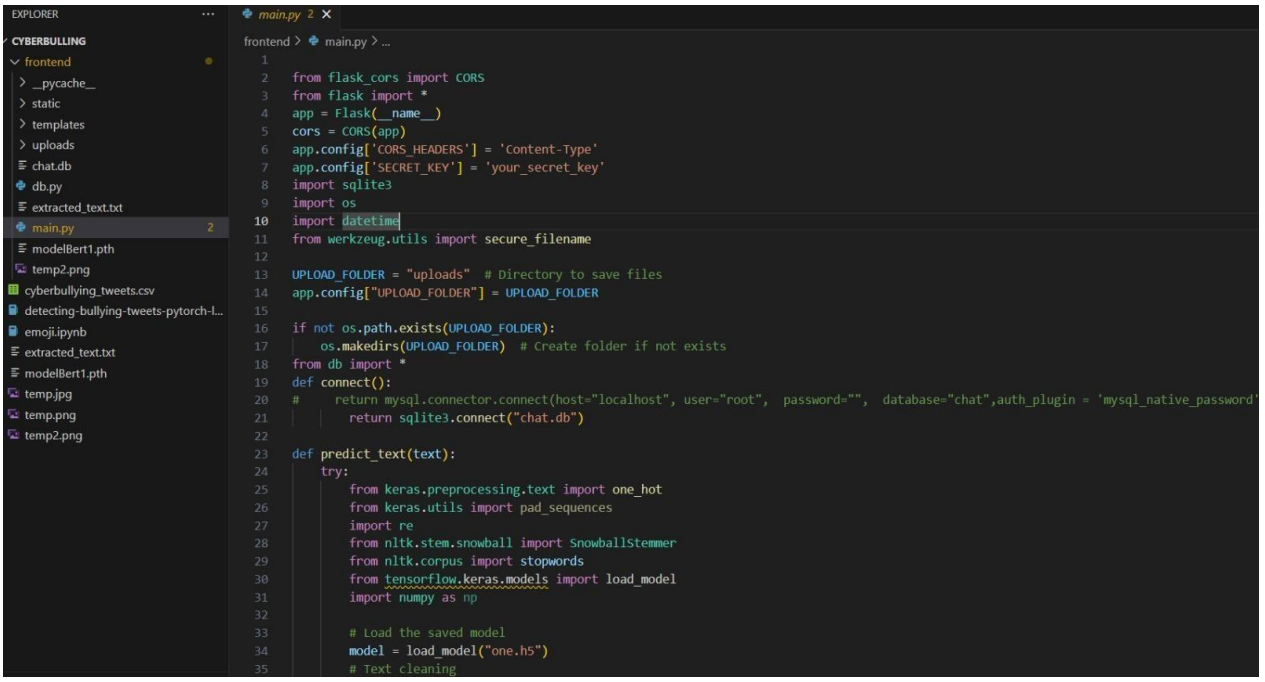
Real-Time Deployment: Implement the system in real-time environments (e.g., integration with social media APIs) for live monitoring and immediate intervention.

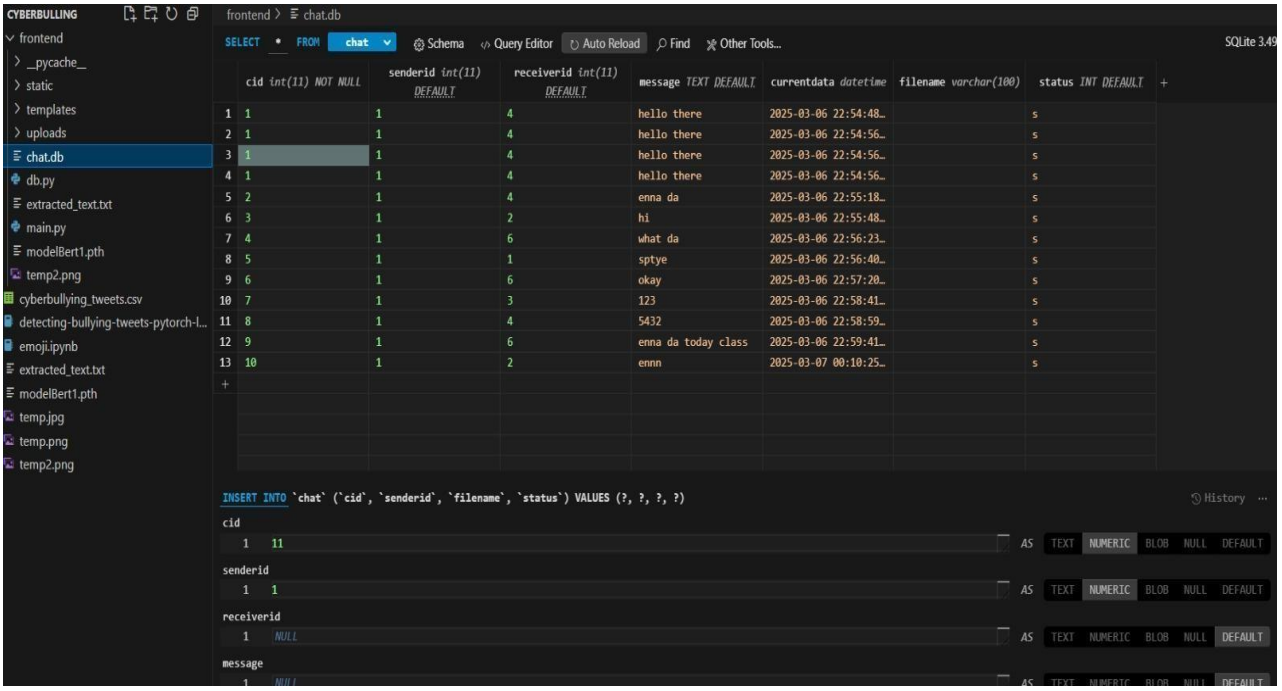
Lightweight Models: Explore using efficient transformer variants like DistilBERT or MobileBERT to reduce computational load and enable deployment on mobile devices.

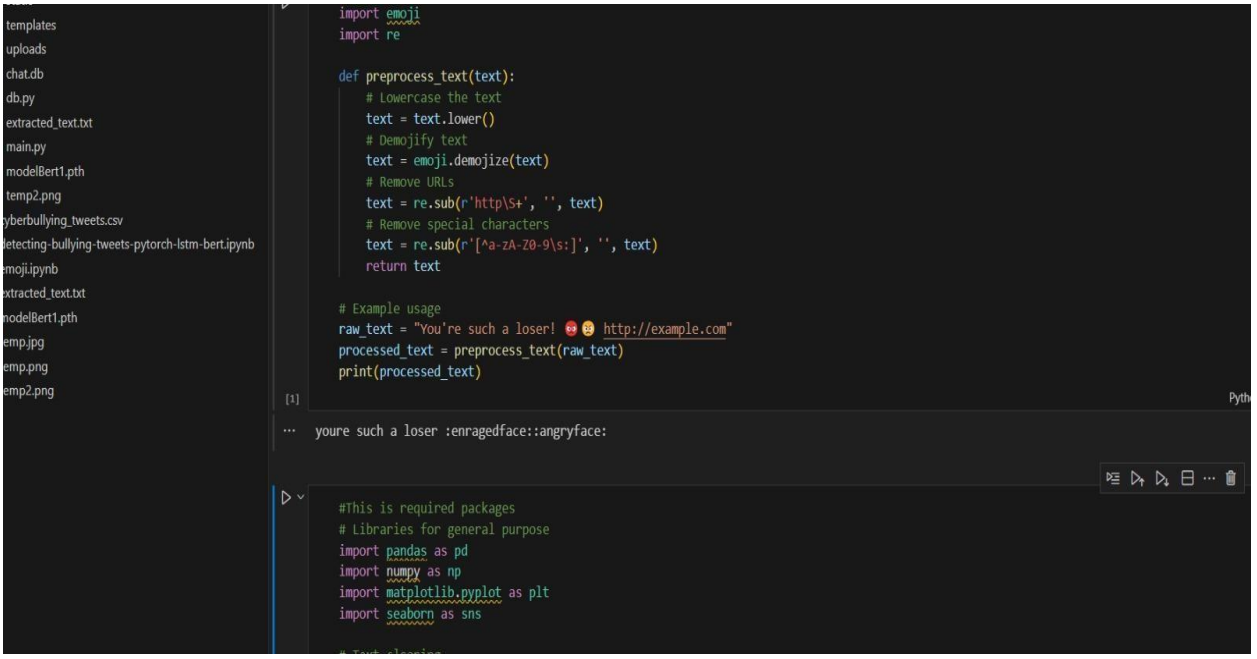
Enhanced Contextual Detection: Incorporate external knowledge bases or social media metadata (e.g., user profiles, reply chains) for deeper contextual analysis.

APPENDIX

APPENDIX 1




CYBERBULLING
frontend
__pycache__
static
templates
uploads
chat.db
db.py
extracted_text.txt
main.py
modelBert1.pth
temp2.png
cyberbullying_tweets.csv
detecting-bullying-tweets-pytorch-l...
emoji.ipynb
extracted_text.txt
modelBert1.pth
temp.jpg
temp.png
temp2.png
frontend > chat.db
SELECT * FROM chat
Schema
Query Editor
Auto Reload
Find
Other Tools...
SQLite 3.49
cid int(11) NOT NULL
senderid int(11) DEFAULT
receiverid int(11) DEFAULT
message TEXT DEFAULT
currentdata datetime
filename varchar(100)
status INT DEFAULT
1 1 1 4 hello there 2025-03-06 22:54:48... s
2 1 1 4 hello there 2025-03-06 22:54:56... s
3 1 1 4 hello there 2025-03-06 22:54:56... s
4 1 1 4 hello there 2025-03-06 22:54:56... s
5 2 1 4 enna da 2025-03-06 22:55:18... s
6 3 1 2 hi 2025-03-06 22:55:48... s
7 4 1 6 what da 2025-03-06 22:56:23... s
8 5 1 1 sptye 2025-03-06 22:56:40... s
9 6 1 6 okay 2025-03-06 22:57:20... s
10 7 1 3 123 2025-03-06 22:58:41... s
11 8 1 4 5432 2025-03-06 22:58:59... s
12 9 1 6 enna da today class 2025-03-06 22:59:41... s
13 10 1 2 ennn 2025-03-07 00:10:25... s
INSERT INTO `chat` (`cid`, `senderid`, `filename`, `status`) VALUES (?, ?, ?, ?)
History
cid
1 11
AS TEXT NUMERIC BLOB NULL DEFAULT
senderid
1 1
AS TEXT NUMERIC BLOB NULL DEFAULT
receiverid
1 NULL
AS TEXT NUMERIC BLOB NULL DEFAULT
message


templates
uploads
chat.db
db.py
extracted_text.txt
main.py
modelBert1.pth
temp2.png
yberbullying_tweets.csv
detecting-bullying-tweets-pytorch-lstm-bert.ipynb
moji.ipynb
xtracted_text.txt
odelBert1.pth
mp.jpg
mp.png
mp2.png
import emoji
import re
def preprocess_text(text):
# Lowercase the text
text = text.lower()
# Demojify text
text = emoji.demojize(text)
# Remove URLs
text = re.sub(r'http\S+', '', text)
# Remove special characters
text = re.sub(r'[^a-zA-Z0-9\s:]', '', text)
return text
# Example usage
raw_text = "You're such a loser! http://example.com"
processed_text = preprocess_text(raw_text)
print(processed_text)
[1]
youre such a loser :enragedface::angryface:
#This is required packages
# Libraries for general purpose
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

APPENDIX 2

CERTIFICATE

OF PARTICIPATION

Global Conference Hub

This is to Certify that

Dr.R.G.Suresh Kumar

Department of Computer Science Engineering, Rajiv Gandhi College of Engineering and Technology, Tamilnadu, India.

For attending and giving an oral presentation on the paper entitled

Cyberbullying detection and prevention using machine learning

at the "Second International Conference on Advancements in Science, Technology and Management (ICASTM-II) 2025 "- Jointly Organized by the Department of Computer Science, St. Xavier's College of Management & Technology (SXCMT), Patna, India & Global Conference Hub, Coimbatore, Tamil Nadu, India on 11th &12th April 2025.

Dr. Supriya Shree
Assistant Professor, Department of Computer Science, St. Xavier's College of Management & Technology, Patna, India.

Dr. Martin Poras SJ
Principal, St. Xavier's College of Management & Technology, Patna, India.

SXCMT & GCH ICASTM 2025-113

CERTIFICATE

OF PARTICIPATION

Global Conference Hub

This is to Certify that

Mrs.Elakia K

Department of Computer Science Engineering, Rajiv Gandhi College of Engineering and Technology, Tamilnadu, India.

For attending and giving an oral presentation on the paper entitled

Cyberbullying detection and prevention using machine learning

at the "Second International Conference on Advancements in Science, Technology and Management (ICASTM-II) 2025 "- Jointly Organized by the Department of Computer Science, St. Xavier's College of Management & Technology (SXCMT), Patna, India & Global Conference Hub, Coimbatore, Tamil Nadu, India on 11th &12th April 2025.

Dr. Supriya Shree
Assistant Professor, Department of Computer Science, St. Xavier's College of Management & Technology, Patna, India.

Dr. Martin Poras SJ
Principal, St. Xavier's College of Management & Technology, Patna, India.

SXCMT & GCH ICASTM 2025-114


CERTIFICATE
OF PARTICIPATION
This is to Certify that
Mr.Dinesh Kumar H
Department of Computer Science Engineering, Rajiv Gandhi College of Engineering and Technology, Tamilnadu, India.
For attending and giving an oral presentation on the paper entitled
Cyberbullying detection and prevention using machine learning
at the "Second International Conference on Advancements in Science, Technology and Management (ICASTM-II) 2025 "- Jointly Organized by the Department of Computer Science, St. Xavier's College of Management & Technology (SXCMT), Patna, India & Global Conference Hub, Coimbatore, Tamil Nadu, India on 11th &12th April 2025.
Global Conference Hub
Dr. Supriya Shree
Assistant Professor, Department of Computer Science, St. Xavier's College of Management & Technology, Patna, India.
Dr. Martin Poras SJ
Principal, St. Xavier's College of Management & Technology, Patna, India.
SXCMT & GCH ICASTM 2025-112


CERTIFICATE
OF PARTICIPATION
This is to Certify that
Mr.Daniyalraj K
Department of Computer Science Engineering, Rajiv Gandhi College of Engineering and Technology, Tamilnadu, India.
For attending and giving an oral presentation on the paper entitled
Cyberbullying detection and prevention using machine learning
at the "Second International Conference on Advancements in Science, Technology and Management (ICASTM-II) 2025 "- Jointly Organized by the Department of Computer Science, St. Xavier's College of Management & Technology (SXCMT), Patna, India & Global Conference Hub, Coimbatore, Tamil Nadu, India on 11th &12th April 2025.
Global Conference Hub
Dr. Supriya Shree
Assistant Professor, Department of Computer Science, St. Xavier's College of Management & Technology, Patna, India.
Dr. Martin Poras SJ
Principal, St. Xavier's College of Management & Technology, Patna, India.
SXCMT & GCH ICASTM 2025-111

CERTIFICATE

OF PARTICIPATION

This is to Certify that

Yogesh P

Department of Computer Science Engineering, Rajiv Gandhi College of Engineering and Technology, Tamilnadu, India.

For attending and giving an oral presentation on the paper entitled

Cyberbullying detection and prevention using machine learning

at the "Second International Conference on Advancements in Science, Technology and Management (ICASTM-II) 2025 "- Jointly Organized by the Department of Computer Science, St. Xavier's College of Management & Technology (SXCMT), Patna, India & Global Conference Hub, Coimbatore, Tamil Nadu, India on 11th &12th April 2025.

Dr. Supriya Shree
Assistant Professor, Department of Computer Science, St. Xavier's College of Management & Technology, Patna, India.

Dr. Martin Poras SJ
Principal, St. Xavier's College of Management & Technology, Patna, India.

SXCMT & GCH ICASTM 2025-115

CERTIFICATE

OF PARTICIPATION

This is to Certify that

Vishva J

Department of Computer Science Engineering, Rajiv Gandhi College of Engineering and Technology, Tamilnadu, India.

For attending and giving an oral presentation on the paper entitled

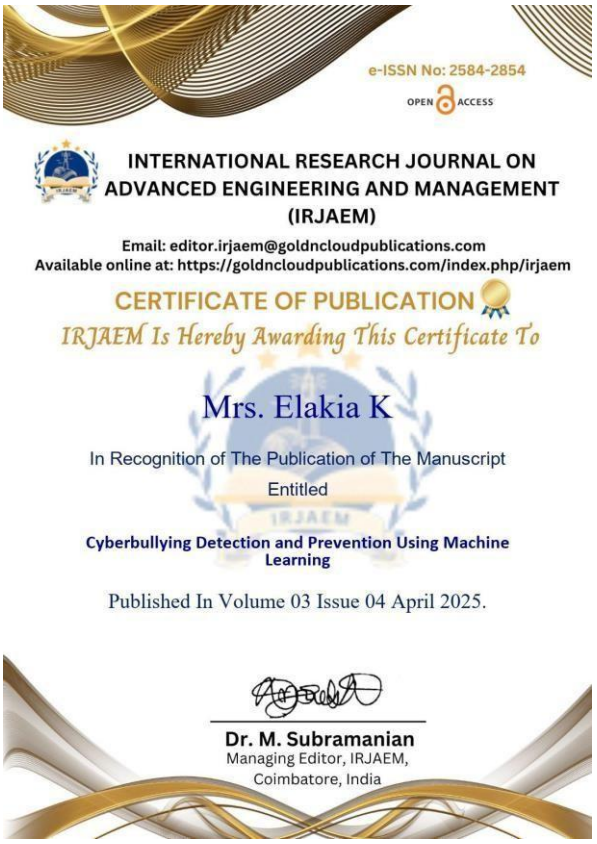
Cyberbullying detection and prevention using machine learning

at the "Second International Conference on Advancements in Science, Technology and Management (ICASTM-II) 2025 "- Jointly Organized by the Department of Computer Science, St. Xavier's College of Management & Technology (SXCMT), Patna, India & Global Conference Hub, Coimbatore, Tamil Nadu, India on 11th &12th April 2025.

Dr. Supriya Shree
Assistant Professor, Department of Computer Science, St. Xavier's College of Management & Technology, Patna, India.

Dr. Martin Poras SJ
Principal, St. Xavier's College of Management & Technology, Patna, India.

SXCMT & GCH ICASTM 2025-116



e-ISSN No: 2584-2854

OPEN ACCESS

INTERNATIONAL RESEARCH JOURNAL ON ADVANCED ENGINEERING AND MANAGEMENT (IRJAEM)

Email: editor.irjaem@goldncloudpublications.com
Available online at: https://goldncloudpublications.com/index.php/irjaem

CERTIFICATE OF PUBLICATION

IRJAEM Is Hereby Awarding This Certificate To

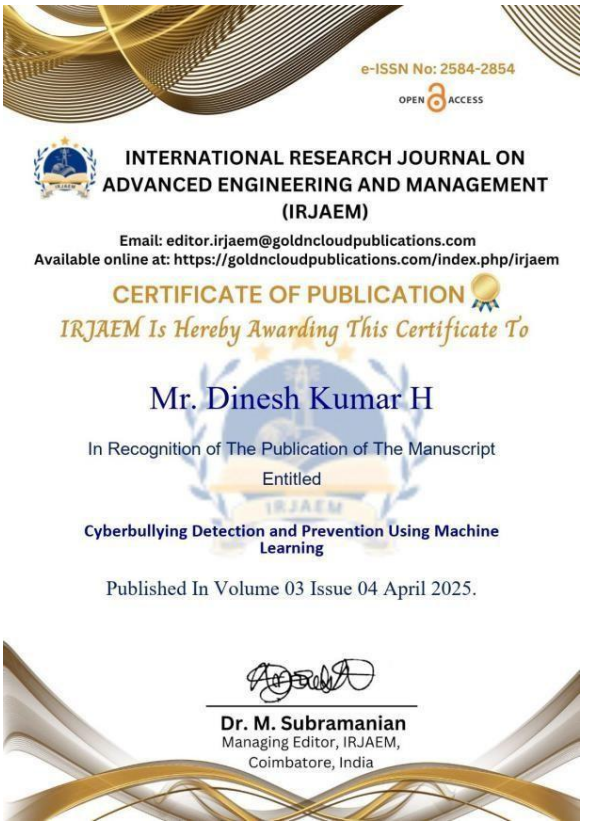
Mrs. Elakia K

In Recognition of The Publication of The Manuscript Entitled

Cyberbullying Detection and Prevention Using Machine Learning

Published In Volume 03 Issue 04 April 2025.

Dr. M. Subramanian
Managing Editor, IRJAEM,
Coimbatore, India



e-ISSN No: 2584-2854

OPEN ACCESS

INTERNATIONAL RESEARCH JOURNAL ON ADVANCED ENGINEERING AND MANAGEMENT (IRJAEM)

Email: editor.irjaem@goldncloudpublications.com
Available online at: https://goldncloudpublications.com/index.php/irjaem

CERTIFICATE OF PUBLICATION

IRJAEM Is Hereby Awarding This Certificate To

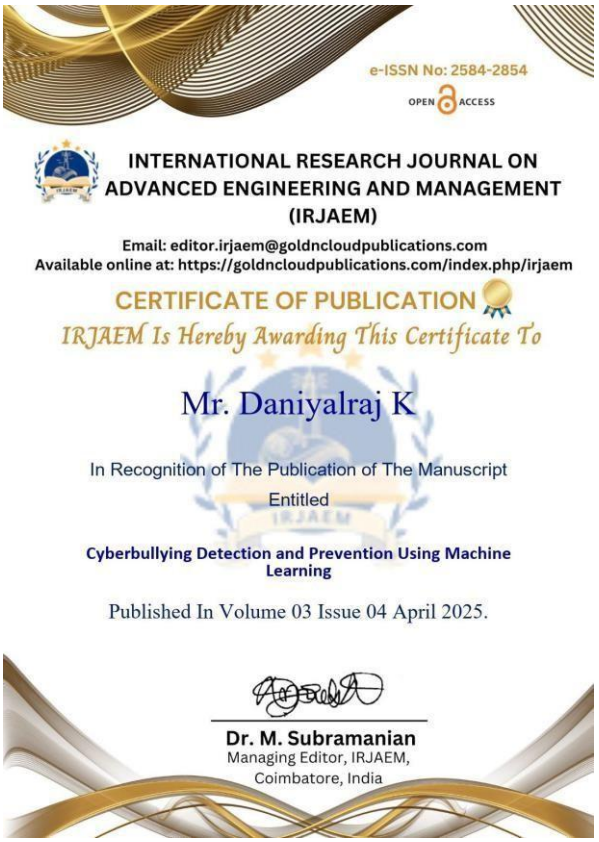
Mr. Dinesh Kumar H

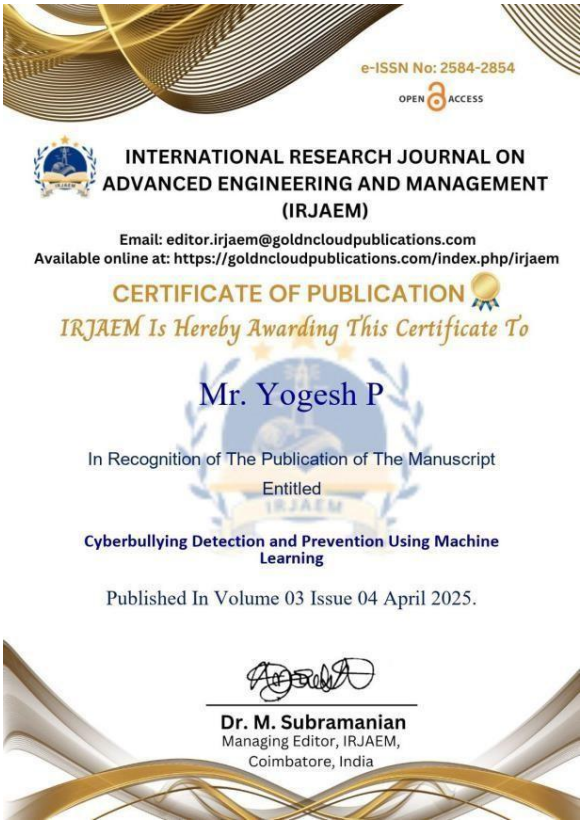
In Recognition of The Publication of The Manuscript Entitled

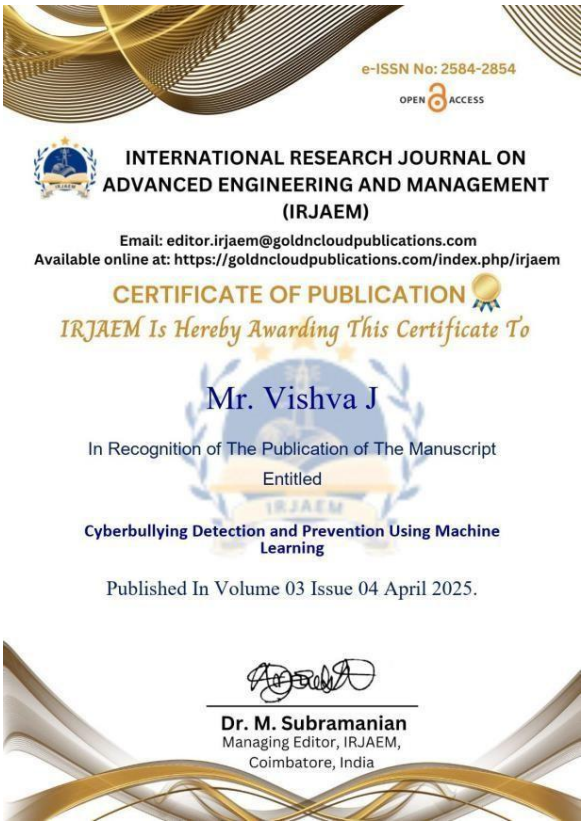
Cyberbullying Detection and Prevention Using Machine Learning

Published In Volume 03 Issue 04 April 2025.

Dr. M. Subramanian
Managing Editor, IRJAEM,
Coimbatore, India


e-ISSN No: 2584-2854
OPEN ACCESS
INTERNATIONAL RESEARCH JOURNAL ON ADVANCED ENGINEERING AND MANAGEMENT (IRJAEM)
Email: editor.irjaem@goldncloudpublications.com
Available online at: https://goldncloudpublications.com/index.php/irjaem
CERTIFICATE OF PUBLICATION
IRJAEM Is Hereby Awarding This Certificate To
Mr. Daniyalraj K
In Recognition of The Publication of The Manuscript
Entitled
Cyberbullying Detection and Prevention Using Machine Learning
Published In Volume 03 Issue 04 April 2025.
Dr. M. Subramanian
Managing Editor, IRJAEM,
Coimbatore, India


e-ISSN No: 2584-2854
OPEN ACCESS
INTERNATIONAL RESEARCH JOURNAL ON ADVANCED ENGINEERING AND MANAGEMENT (IRJAEM)
Email: editor.irjaem@goldncloudpublications.com
Available online at: https://goldncloudpublications.com/index.php/irjaem
CERTIFICATE OF PUBLICATION
IRJAEM Is Hereby Awarding This Certificate To
Mr. Yogesh P
In Recognition of The Publication of The Manuscript
Entitled
Cyberbullying Detection and Prevention Using Machine Learning
Published In Volume 03 Issue 04 April 2025.
Dr. M. Subramanian
Managing Editor, IRJAEM,
Coimbatore, India



e-ISSN No: 2584-2854

OPEN ACCESS

INTERNATIONAL RESEARCH JOURNAL ON ADVANCED ENGINEERING AND MANAGEMENT (IRJAEM)

Email: editor.irjaem@goldncloudpublications.com
Available online at: https://goldncloudpublications.com/index.php/irjaem

CERTIFICATE OF PUBLICATION

IRJAEM Is Hereby Awarding This Certificate To

Mr. Vishva J

In Recognition of The Publication of The Manuscript Entitled

Cyberbullying Detection and Prevention Using Machine Learning

Published In Volume 03 Issue 04 April 2025.

Dr. M. Subramanian
Managing Editor, IRJAEM,
Coimbatore, India

REFERENCE

[1] Amshuman Singh-―Machine Learning Approach to Crime Prediction and Identification of Hotspots," Published: August 2021.

[2] Neil Shah, Nandish Bhagat & Manan Shah- ―Crime forecasting: a machine learning and computer vision approach to crime prediction and prevention‖, Published: April 2021.

[3] Md Manowarul Islam, Md Ashraf Uddin, Linta Islam- ―Cyberbullying Detection on Social Networks Using Machine Learning Approaches‖, Published:2020.

[4] Bandeh Ali Talpur, Declan O'Sullivan- ―Cyberbullying severity detection: A machine learning approach‖ Published: October 2020.

[5] Department of Translation, Interpreting, and Communication - Faculty of Arts and Philosophy, Ghent University, Ghent, Belgium, ―Automatic detection of cyberbullying in a social media text.‖ Published online 2018 Oct 8.

[6] John Hani, Mohamed Nashaat, Mostafa Ahmed, Zeyad Emad, Eslam Amer, Ammar Mohammed, ―Social Media Cyberbullying Detection using Machine Learning‖. An Article Published in an International Journal of Advanced Computer Science and Applications(IJACSA), Volume 10 Issue 5, 2019.

[7] Nureni Ayofe Azeez, Sunday O. Idiakose, Chinazo Juliet Onyema and Charles Van Der Vyver, ―Cyberbullying Detection in Social Networks: Artificial Intelligence Approach‖ Publication 18 June 2021.

[8] Afrah Almansoori, Mohammed Alshamsi, Sherief Abdallah, and Said A. Salloum, ―Analysis of Cybercrime on Social Media Platforms and Its Challenges‖ under exclusive licence to Springer Nature Switzerland AG 2021.

[9] Kazi Saeed Alam, Shovan Bhowmik, Priyo Ranjan Kundu Prosun- ―Cyberbullying Detection: An Ensemble Based Machine Learning Approach," Published: July 2021.

[10] A.Guazzelli, M. Zeller, W. Chen, and G. Williams, ―PMML an open standard for sharing models,‖ R J., vol. 1, no. 1, pp. 60–65, May 2009.

[11] M. Hall, E. Frank, J. Holmes, B. Pfahringer, P. Reutemann, and I. Witten, ―The WEKA data mining software: An update,‖ ACM SIGKDD Explor. Newsletter., vol. 11, no. 1, pp. 10–18, 2009.

[12] R Language Definition. (2000). R Core Team [Online]. Available: ftp://155.232.191.133/cran/doc/manuals/r-devel/R-lang.pdf, accessed on Nov. 2015.

[13] M.Graczyk, T.Lasota, and B.Trawinski, ―Comparative analysis of premises valuation models using KEEL, RapidMiner, and WEKA,‖ Computational Collective Intelligence. Semantic Web, Social Networks and Multiagent Systems. New York, NY, USA: Springer, 2009.

[14] V. Jacobson, C. Leres, and S. McCanne, The Tcpdump Manual Page. Berkeley, CA, USA: Lawrence Berkeley Laboratory, 1989.

[15] G. F. Lyon, Nmap Network Scanning: The Official Nmap Project Guide to Network Discovery and Security Scanning." USA: Insecure, 2009.

[16] R. Lippmann, J. Haines, D. Fried, J. Korba, and K. Das, ―The 1999 DARPA offline intrusion detection evaluation,‖ Comput. Netw., vol. 34,pp. 579–595, 2000.

[17] R. Lippmann et al., ―Evaluating intrusion detection systems: The 1998 DARPA off-line intrusion detection evaluation,‖ in Proc. IEEE DARPAInf. Survive. Conf. Expo., 2000.

[18] S. J. Stolfo, KDD Cup 1999 Data Set, University of California Irvine, KDD repository [Online]. Available: http://kdd.ics.uci.edu, accessed on Jun. 2014.

[19] S. J. Stolfo, KDD Cup 1999 Data Set, University of California Irvine, KDD repository [Online]. Available: http://kdd.ics.uci.edu, accessed on Jun. 2014.

[20] Aizenkot D, Kashy-Rosenbaum G (2018) Cyberbullying in WhatsApp classmates' groups: evaluation of an intervention program implemented in israeli elementary and middle schools. New Media & Society 20(12):4709–4727

[21] Aldhyani TH, Al-Adhaileh MH, Alsubari SN (2022) Cyberbullying identification system based deep learning algorithms. Electronics 11(20):3273

[22] Al-Garadi MA, Hussain MR, Khan N, Murtaza G, Nweke HF, Ali I, …, Gani A (2019) Predicting cyberbullying on social media in the big data era using machine learning algorithms: review of literature and open challenges. IEEE Access 7:70701–70718

[23] Kumar A, Sachdeva N (2020) Multi-input integrative learning using deep neural networks and transfer learning for cyberbullying detection in real-time code-mix data. Multimedia Systems

[24] Monteiro RP, Santana MC, Santos RM, Pereira FC (2022) Cyberbullying victimization and mental health in higher education students: the mediating role of perceived social support. J interpers Violence, 1–23

[25] Nandhini BS, Sheeba JI (2015) Online social network bullying detection using intelligence techniques. Procedia Comput Sci 45:485–492

[26] Shah N, Maqbool A, Abbasi AF (2021) Predictive modeling for cyberbullying detection in social media. J Ambient Intell Humaniz Comput 12(6):5579–5594

[27] Soni D, Singh VK (2018) See no evil, hear no evil: Audio-visual-textual cyberbullying detection. Proceedings of the ACM on Human-Computer Interaction, 2(CSCW), 1–26

[28] Van Hee C, Lefever E, Verhoeven B, Mennes J, Desmet B, De Pauw G, …, Hoste V (2015) Detection and fine-grained classification of cyberbullying events. In International Conference Recent Advances in Natural Language Processing (RANLP).

[29] Zhao R, Zhou A, Mao K (2016) Automatic detection of cyberbullying on social networks based on bullying features. In Proceedings of the 17th international conference on distributed computing and networking.