International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 04 April 2025
Page No: 2007-2015
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0293

BERT-Based Cyberbullying Detection

Maladhy D[1], Jeevitha D[2], Madhumitha K[3], Subashree J[4]
[1]Assistant Professor, Dept. of IT, Rajiv Gandhi College of Engg. & Tech., Kirumampakkam, Puducherry, India.
[2,3,4]UG Scholar, Dept. of IT, Rajiv Gandhi College of Engg. & Tech., Kirumampakkam, Puducherry, India.
Emails: *dmaladhy_it@rgcet.edu.in[1], jeevithaduraimurugan@gmail.com[2], madhumithakanda2001@gmail.com[3], subekavi03@gmail.com[4]*

Abstract
Cyberbullying on social media has been on the rise lately-and with it, the serious psychological effects it leaves in its wake: anxiety and sadness. That's why early detection and intervention are so crucial. Those traditional methods of tackling online abuse often fall short when it comes to slang and ever-changing language. That's because they just can't pick up on the intent behind the words. Our project tackles that problem head-on by combining text and visual elements in a way that deep learning can really understand. We use a refined BERT model to put language in context, Demoji to decipher the meanings of emojis and Pytesseract to extract text from images. That hybrid approach ensures even the most hidden or indirect bullying messages are identified. We deliver that analysis-and the tools to visualize it-in real-time through a mobile app. That means non-technical users-parents, teachers and moderators-can easily use it to spot and stop cyberbullying. By harnessing the latest AI technologies to safeguard vulnerable people, we create a safer online environment.

Keywords: *Cyberbullying, Real-time Analysis, BERT Model, Demoji, Pytesseract.*

1. Introduction
The world has evolved through different dimensions through internet in various fields like education, sports, entertainment etc,. As there are ups and downs in life, the internet also has its own downsides. The biggest problem in this digital world is cyberbullying. In recent surveys it has been shown that about 36.5% of the respondents have gone through cyberbullying, by facing harassment through digital media. Increase in internet usage lead to increase in cyberbullying activities at an alarming rate. 87% of the young social media users have accepted that they have been through these kinds of online harassments. It's a tough task to deal with cyberbullying as it can happen in different ways like toxic comments, photographs, videos etc,. Advanced technologies have been used to detect and remove cyberbullying activities on social platforms to prevent these kinds of harmful happenings. But cyberbullying detection is a very tough task. Sometimes when we are having a normal conversation with friends, they might sound like bullying but if we examine properly it might not be the same. Many studies based on the Cyberbullying detection suggested the application of traditional machine learning-based models and advanced deep learning-based models for better accuracy. The deep learning-based models used GloVe and SSWE for different word embedding techniques. The results conclude that more than those generated from traditional machine learning models, deep learning-based models work better consistently. A new approach for cyberbullying detection is developed by using a pre-trained BERT that outperform on numerous NLP tasks. Such models are applicable in catching contextual meanings of words and phrases, thereby able to dig deeper into the complexities of online communication. Research says that the deep learning-based models would beat the traditional models when it comes to detecting cyberbullying tasks. Using this advancement, there came BERT-based detection, which is actually an innovation developed by Google AI that marks an enormous leap in the world of natural language processing. Also, BERT can be fine-tuned over specific tasks, which makes it fit perfectly to handle the task of harmful social media content identification. Though Cyberbullying is a serious and growing problem but the development of deep learning models, specifically BERT, might bring some hope for

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 04 April 2025
Page No: 2007-2015
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0293

solutions.

2. Literature Survey

Cyber dating abuse in adolescents Myths of romantic love, sexting practices and bullying [1] Ainize Martínez Soto, Cristina Lopez-del Burgo [1] This study shows the complexities of cyber dating abuse (CDA) in teens and we need programs that cater to different social factors. Through a diverse sample, we found boys are more likely to bully, sext and believe in romantic love myths, while highly religious teens are less likely to sext. So we need to educate teens on safe online behavior and identify abusive actions in virtual and real life relationships. We can better protect them by considering the different influences of gender, culture and personal beliefs on their online interactions.

Cyberbullying Detection in Social Networks: A Comparison Between Machine Learning and Transfer Learning Approaches [2] Ainize Martínez Soto , Cristina Lopez-del Burgo [2] The research developed an automatic system for detecting cyberbullying using two distinct approaches: Conventional Machine Learning (CML) and Transfer Learning. The CML approach, which utilized a combination of textual, sentiment, emotional, and toxicity features within a Logistic Regression model, achieved an F-measure of 64.8%. Among the various embeddings evaluated, DistilBert emerged as the most effective, yielding the highest F-measure both individually and in combination with other features.

A robust hybrid machine learning model for Bengali cyber bullying detection in social media [3] Arnisha Akhter , Uzzal Kumar Acharjee , Md. Alamin Talukder [3] This article introduces a new hybrid machine learning approach that centers on detecting cyberbullying in Bengali on social media platforms. The study uses advanced text preprocessing methodologies and adopts TfidfVectorizer of effective feature representation. The method outperforms prior methods of detectable Bengali cyberbullying and provides substantial contributions to better protection of users on social media.

A Review on Deep-Learning-Based Cyberbullying Detection [4] Md. Tarek Hasan, Md. Al Emran Hossain, Md. Saddam Hossain Mukta [4] This paper emphasizes the advantages of deep learning models over conventional machine learning algorithms regarding cyberbullying detection. Deep learning models utilize their data processing capacity and data classification abilities to classify images and text with a better degree of automating feature extraction through hidden layers. The result is more accurate pattern identification and cyberbully detection when compared to conventional algorithms. This paper also reviews datasets and deep learning architectures proposed in the studies mentioned, describing the tasks carried out each time a dataset was used. Ultimately, this paper not only adds to the understanding of present methods, but sets the stage for future research using deep learning and other methods to improve cyberbullying detection systems.

Detecting cyberbullying using deep learning techniques: a pre-trained glove and focal loss technique [5] Amr Mohamed El Koshiry , Entesar Hamed I. Eliwa , Tarek Abd El-Hafeez [5] This research comprises a performance comparison, using appropriate metrics (i.e., accuracy, precision, recall, and F1 score) among several classical and deep learning algorithms to detect the presence of cyberbullying in tweets. The Focal Loss algorithm exhibited the strongest performance based on accuracy and precision metrics; however, the low recall across most algorithms suggests a challenge in detecting all instances of cyberbullying among the tweets. This research proposes novel directions for detection research and application by integrating a convolutional neural network (CNN) with a bidirectional long short-term memory (Bi-LSTM) layer. The detection model is trained on a pre-processed dataset of tweets and is trained on GloVe word embeddings and the Focal Loss function to increase accuracy in detecting any instances of cyberbullying.

3. Architecture Diagram

The recommended structure for the cyberbullying detection system allows for efficient data processing, real-time detection, and precision across the various formats. This architecture is made up of three layers: User Interface Layer, Processing Layer, and Backend & Database Layer. The users' interactions with the system take place via a web or mobile application, using React Native, that allows the user to submit text, images, or emojis for processing. The Processing Layer uses a highly tuned BERT model to perform deep processing of the text, Pytesseract will extract text from any images submitted by the user, and Demoji will interpret emojis, allowing for the detection of bullying communicated in all three formats. The Backend & Database Layer uses Python with Flask for processing of the user request and stores the information captured from the users into a structured SQL/NoSQL tied to further processing. The proposed architecture allows for user data to be processed efficiently, securely and accurately, which will support a reliable and efficient method for detecting instances of cyberbullying, Figure 1.

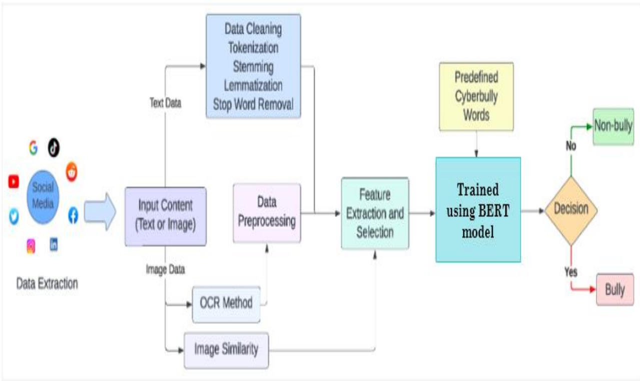

Figure 1 **Architecture Diagram**

4. 3. Use Case Diagram:

The diagram of use case for the integrated cyberbullying detection plan explains how the different actors interact with the different capabilities of the plan. The main actor of this plan will be the user, although in reality a person communicates via social media as the user. The user of this plan interacts with the plan by posting, commenting or sending a message to someone, which may involve a form of being subjected to cyberbullying. Anytime a post of any kind is sent, the plan will process the post which may involve data collection, data preprocessing, extracting features, and classifying data. The overall purpose of the plan is to identify and alert the user in cases of language and content that is harmful to other people. The diagram illustrates the proposed overall structure of the plan to show the interactions of the actors and the subcomponents of the use case which highlights how user interaction will provide the necessary action to generate proactive or intervention strategies with regard to cyberbullying while contributing to inclusive and safe online experiences, shown in Figure 2.

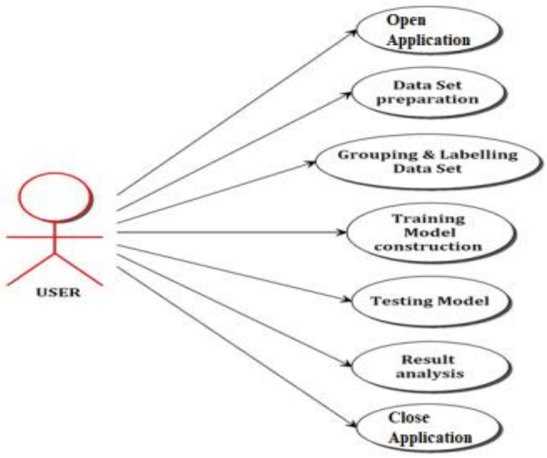

Figure 2 **Use Case Diagram**

5. Sequence Diagram

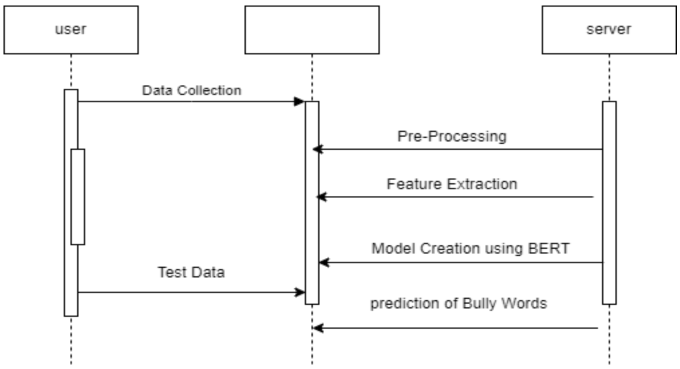

Figure 3 **Sequence Diagram**

This sequence diagram illustrates the interactions between user, system, and components of the system throughout the detection process step-by-step. Initially, a user will post or comment on the social media platform. At this point, the system recognizes the submission, kicking off a sequence of events starting with data gathering. After that, the system follows-up with pre-processing steps to the submitted text in order to clean the text and prepare it for processing. After pre-processing, the system sends the cleaned text to a BERT model for feature extraction, which generates numeric values representing the content in the context of meaning. After my BERT model features, the model is then trained to classify the user submission as bullying or non-bullying. In the end, the system returns the classification back to the users along with alerts or actions if the content contained any harmful material. Overall, this sequence of observations highlights the structure and flow of data and processing within the system and highlights the organizations process flow aligns to user engagement, shown in Figure 3.

6. Architecture Diagram of The Bert Model

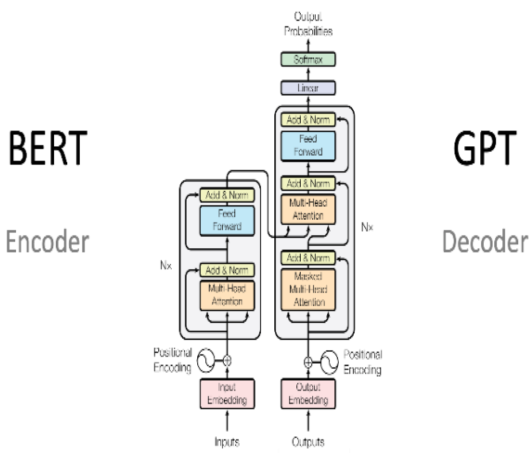

Figure 4 BERT Architecture Diagram

BERT (Bidirectional Encoder Representations from Transformers) model is based on the architecture of transformers which is particularly constructed for language understanding. BERT model is composed of numerous layers of transformers which use self-attention to identify the relevance or relationship between words in a sentence, irrespective of their location in the sentence. The model utilizes this self-attention model framework in which context is provided bidirectionally from the left and right side of each token within a sentence to produce a more natural understanding of language. When fine-tuned to a specific task such as bullying language detection, BERT is able to apply its wide-ranging context-based knowledge in predicting based on textual input. In summary, the architecture diagram depicts the interconnected components, including how data flows and between these components and the various layers of the architecture all designed to assist the understanding and generative capabilities of the Bert decision framework, shown in Figure 4.

7. Proposed System

The cyberbullying detection system has been enhanced using deep learning and natural language processing (NLP) methods to evaluate text, emojis, and images. It takes advantage of a fine-tuned BERT model to enhance the understanding of language by determining the contextual meaning of a sentence, rather than of single words. Pytesseract is to be utilized to extract text that could be hidden in the image for a greater detection of textual content capacity so that it can evaluate images that are often used to bully others, such as memes and screenshots. Demoji will be used to interpret the meaning of emojis so that any meaning or sarcasm conveyed in the symbols are not ignored. Overall, the system is using these technologies to improve the ability to detect cyberbullying both more accurately while at the same time decreasing its risk of a false negative and improving the reliability overall [6-8].

7.1 Data Collection

Data collection is a vital stage in building a more precise cyberbullying detection system, involving the gathering of different datasets from a variety of digital spaces, including text, images, and video. The goal of the data collection is to create a holistic dataset that carries representations of multiple forms of online bullying, so that the model can recognize harmful interactions across different contexts. The datasets we are collecting contain chat conversations, comments, Twitter posts, and images that are typically platforms where cyberbullying occurs. One or more forms of multimodal content sample will also be collected. Cyberbullying occurs primarily in text, however, often includes images, memes, screenshots, and emojis multi-modal content, which if we only use text we will inadvertently miss. This will allow us to identify indirect, covert, or sarcastic forms of bullying, that are not identifiable using typical text-based approaches. All personally identifiable information (PII) will be removed or anonymized for compliance with data protection frameworks like GDPR and CCPA. The identified sample will be auto-filtered, then manually filtered to remove irrelevant samples that may be biased to provide, as a goal, a more balanced dataset that represents a sample of real-world cyberbullying incidents. This provides the detection system with the ability to recognize different types of online harassment, and improve detection, intervention, and prevention strategies in a more effective way.

7.2 Pre-processing

Pre-processing utilizes raw data for use in a data analysis technique by standardizing and cleaning it. For example, for text data pre-processing will entail removing stopwords, punctuation, special characters, and other irrelevant aspects which do not provide substantive analysis. Pre-processing may also entail what we refer to the process of cleaning up the variations of the words, as you will see later on. This entails converting the words to lowercase, stemming, and lemmatizing (or normalizing) the variations of words to make sure they have the same general representation. This cleaning up process is particularly relevant for slang, abbreviations, and other informal expressions which occur very frequently in online communication. For image data pre-processing takes place to improve the accuracy of Optical Character Recognition (OCR) by resizing and adjusting the contrast of images with text (e.g., memes or screenshots). This is due to the fact that changes will improve OCR efficiency for extracting the embedded text and make it easier for machine-learning models to analyze. Pre-processing of images may also include emojis which may be converted to text descriptions using Demoji, for example, and thus maintain the emotional context that emojis invariably alter in meaning of usages of the words. Consequently, pre-processing develops a structured and standardization process by which the cyberbullying detection model can process the data used for learning analytical submission.

7.3 Feature Extraction

Feature extraction detects patterns related to cyberbullying through the analysis of textual as well as visual information. Text analysis involves key features such as word frequency, n-grams (bigrams, trigrams), sentiment scores, and TF-IDF (Term Frequency-Inverse Document Frequency). These features facilitate the ability of the model to identify hostile language, insults and threats of violence. Sentiment analysis captures negative emotions such as anger, hatred or fear, all of which are often indicators of cyberbullying. For image-based detection, feature extraction extracts text and patterns in images (utilizing OCR) that may suggest malicious intent. The integration of both linguistic and image-based information creates better accuracy for classification and richness of context increasing the ability to detect subtle, indirect or multimodal forms of cyberbullying.

7.4 OCR Module

Pytesseract is an OCR tool that can pull text from images such as screenshots, memes, and social media posts. Cyberbullying may not always be communicated through descriptive text, and therefore analyzing content that utilizes images is important for complete detection. Pytesseract converts the text in an image to a machine-readable format and allows the system to analyze and process visual media, in conjunction with typical text-based information. After text extraction, the text is then processed using NLP and eventually sentiment analysis so that threats, insults, and harmful messaging can be classified correctly. The use of OCR for detecting cyberbullying has the added benefit of conducting multimodal analysis for the system, which may improve the detection of abusive content regardless of the format, providing a safer and more equitable online space across diverse digital formats.

7.5 Emoji Transcription

In cyberbullying, emojis can heighten and reinforce a negative intent, or disguise an abusive message that is often difficult to detect with legacy text-based approaches. Demoji transcribes emojis to text descriptions to allow the system to interpret meaning and intent accurately. For example, an insult followed by a laughing emoji is representative of mockery of the target, whereas another "stable" message, combined with an angry emoji, may represent aggression. Systematically converting emojis to text enables the system to analyze sentiment, intent, and emotional signals in a conversational exchange with greater precision.

7.6 Text Analysis

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 04 April 2025
Page No: 2007-2015
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0293

A finetuned BERT model improves cyberbullying detection by evaluating text in a bidirectional manner, as compared to typically unidirectional models. This means it evaluates each word based on the words that came before it and the words that will come after it. Because it is bidirectional, the model can evaluate context, nuance, and sentiment, meaning it can detect instances of bullying that are more complex such as indirect, sarcastic, or ambiguous types or forms of bullying. A unidirectional model would evaluate the words and relationships one at a time, whereas, BERT evaluates relationships within the individual sentence at a deeper level. Which will lead to detecting examples like, "Great job, genius!" One could easily interpret, "Great job!" as positive, however, the "genius!" gives the impression that the person was being sarcastic. This is apparent due to the context of how the sentence is constructed. By finetuning BERT on a dataset with cyberbullying content, BERT is highly accurate and identifies patterns of abuse with very good recall.

7.7 Test Data

Test data is a distinct set of data that is leveraged to evaluate the performance of a machine learning model and the model's ability to generalize. While the training data is the data that the model learns from, the test data evaluates whether a model can perform in an area the model has not seen. Test data is an important evaluation for performance metrics such as accuracy, precision, recall, and F1 score as it ensures the model punctually identifies instances of cyberbullying as efficiently as possible minimizing false positives and false negatives. Testing different examples in the real-world will also allow for the evaluation of bias, weaknesses, and overfitting which improves the model before and over time deploying the model. Future evaluations and new test data will be analyzed to allow the model to continue to adapt the model to changes and patterns of behavior in cyberbullying and maintain the model's relevance and usefulness into the future.

7.8 Prediction

In this stage, the model that has been previously trained is applied to new data to subsequently predict possible instances of cyberbullying. The system uses deep learning techniques to scope the text, images, and emojis to have a multimodal monitoring approach for detection. In contrast to traditional methods that only analyze the text, this new integrated system allows for a more whole encompassing assessment of online interactions. More specifically, the BERT model is used as a way to identify the text's context, the Pytesseract recognizes the text from the images, and the Demoji tools receives emojis in order to identify a potential bullying intent. Based on combining these models, the aim is to ensure higher detection accuracy, even if the data contains sarcasm, slang, or multimedia. The real-time prediction promises to allow moderators, educators, and other relevant actors to be proactive and take action, aiming to create a safer online space.

8. Result and Discussion

The updated cyberbullying detection system has a superior level of accuracy and reliability, thanks to new deep learning and natural language processing (NLP) tools. While existing text-based methods have benefited from support tools like Benenson's 6-form text categorization system relying on algorithms and keywords, this system can analyze information across various formats since it is multimodal. The ability to analyze images, emojis, and text improves the detection of subtle and indirect forms of cyberbullying because cyberbullying does not have to be straightforward and direct. The modified BERT model considers contextual meanings about language to vastly improve comprehension, providing additional support for the detection of sarcasm, slang, or indirect threats, which reduces the number of false negatives in text-based detection analysis. The OCR program Pytesseract will allow us to extract text from memes and screens, ensuring that any bullying language hidden in the text-as images can be analyzed. The system will use Demoji to allow hidden meanings associated with emojis that may include mockery, sarcasm, or emotional cues. This evidence demonstrates the opportunity provided through the project to promote a safer digital environment through improved detection of cyberbullying in real-world online communications.

8.1 Precision

In the context of detecting cyberbullying using our

IRJAEH
International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 04 April 2025
Page No: 2007-2015
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0293

BERT model, precision matters for evaluating how well the model detected incidents of abusive language (or abusive behavior) in online language. More specifically, since the BERT model detects text as bullying vs. non-bullying, precision is calculated as the ratio of bullying incidents detected (true positives, TP) to the total of detected bullying (true positives, TP) and non-bullying that were incorrectly detected as bullying (false positives (FP)).

$$\text{Precision} = \frac{TP}{TP + FP}$$

- **True Positives (TP):** The number of instances correctly predicted as positive.
- **False Positives (FP):** The number of instances incorrectly predicted as positive.

A high precision metric, in this case, indicates that the model correctly identifies language indicating cyberbullying, or in other words, most of the content it flags as being bullying is in fact bullying. Thus, we reduce false positive content since most of what is flagged is highly true. The representation of a precision metric reinforces the integrity of the detection system we are measuring, allowing us to feel confident that the system is effectively functioning to identify harmful interactions in online environments.

8.2 Recall

Recall is a key metric for measuring the performance of a model and tells us whether the model identified every applicable incident of abusive language or behavior in online content for cyberbullying classification. It is calculated using true positive (TP) divided by true positive (TP) plus false negative (FN), where bullying was present and was not identified by the model.

$$\text{Recall} = \frac{TP}{TP + FN}$$

- **False Negatives (FN):** The number of instances that are actually positive but were incorrectly predicted as negative.

For instance, if BERT is used to classify text messages or social media posts, a high recall score means the model identified a lot of actual abusive comments - a critical factor in terms of prompt response and preventing further abuse. Therefore, recall is often analysed in conjunction with precision - to consider both factors to understand model performance (high recall means inclusion of cases that matter, and minimizing false positive/identifying abusive conversation is high precision).

8.3 F1 Score

The F1 score evaluates a BERT model's performance in cyberbullying detection because it consolidates precision and recall to assess a model's effectiveness in a single evaluation metric. In cyberbullying detection, precision is how many instances the model flagged as bullying were actually abusive, while recall refers to the number of actual cases of cyberbullying the model detected. F1 score is a balanced evaluation of a model's performance especially when there is a trade-off like in precision and recall.

The F1 score is the harmonic mean of precision and recall, calculated as:

$$\text{F1 score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

A model can have a high precision by identifying only a few clear cases of cyberbullying, but missing completely. This would reflect a very low recall rate. This would yield a lower F1 score, which would imply that the model is precise but that there is little overall contribution to the detection of cyberbullying. A similar comment would relate to a case with a high recall for detecting a cyberbullying, but that would mean that the model would detect too many items, and the items might be those that are non-relevant or benign commentary; this would yield a lower precision, and lower F1 score. Ultimately, a high F1 score, in the context of cyberbullying detection, would indicate that the BERT model is capturing a strong proportion of true bullying cases while filtering out non-relevant material, Figure 5.

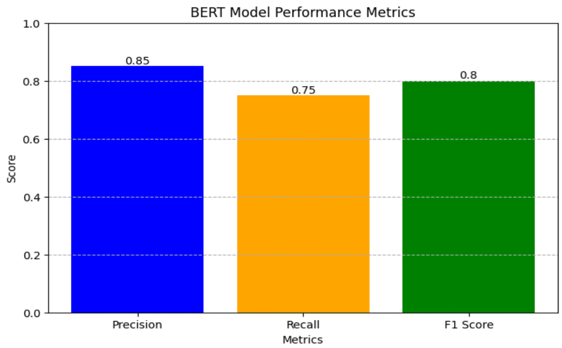

Figure 5 **BERT Model Performance Metrics**

8.4 Accuracy

Accuracy signifies the ratio of actual instances of cyberbullying acknowledged among the overall instances that the model evaluated. In other words, it conveys how often the model predicted whether a comment or message was bullying (positive predictions) or not (negative predictions).

$$\text{Accuracy} = \frac{\text{True Positives (TP)} + \text{True Negatives (TN)}}{\text{Total Predictions (TP + TN + FP + FN)}}$$

where:

- **True Positives (TP)** are sentences correctly identified as important.
- **True Negatives (TN)** are sentences correctly identified as unimportant.
- **False Positives (FP)** are sentences incorrectly identified as important.
- **False Negatives (FN)** are important sentences that were missed by the model.

A high accuracy shows that the model is achieving the desired effect of distinguishing abusive and non-abusive instances across a proportion of instances. Furthermore, when working with a BERT model to classify instances of cyberbullying, a pre-trained BERT model is regularly brought in via packages, such as Hugging Face Transformers, shown in Figure 6.

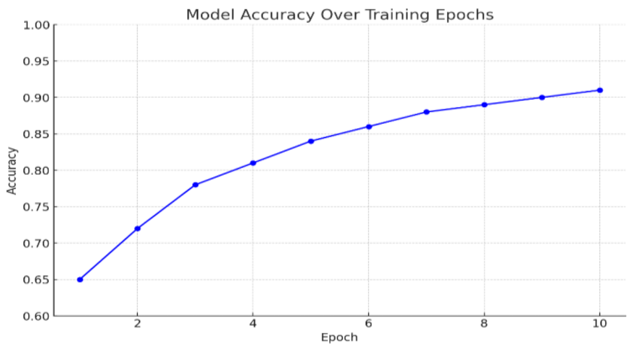

Figure 6 **Model Accuracy Over Training Epochs**

Upon data readiness, it goes through the BERT model that generates embeddings that represent the input's meaning semantically. The output will subsequently undergo processing to classify instances as either cyberbullying or not. By comparing the labeled dataset to the predictions made by the model we will be able to evaluate the model's accuracy identifying strengths and weaknesses. Ultimately, accuracy is a primary indicator of the models reliability and ability to keep users safe from cyberbullying.

Conclusion and Future Work

The earlier phase of the cyberbullying detection system functions as a solid foundation for a more integrated solution for detecting risky online behavior. Enhanced by OCR before images are stored, easy access to emoji text transcription, and the abilities of pretrained BERT for sentiment analysis, the system can process both text and visual data to identify early risk indicators for cyberbullying. In sum, the first functionalities are a valuable stepping stone towards the comprehensive and robust cyberbullying detection, and as the tool iterates, it can potentially emerge as a strong tool not just for capturing cyberbullying as it happens, but also to help create safer online spaces by detecting harmful interactions across a variety of content.

Future work could investigate leveraging the models to extend the capabilities of the model to multiple languages, incorporating the models into social media applications as a real-time system, providing context-based assessments as well as making the task of reporting incidents of cyberbullying less of a cumbersome for users through the reporting interface. Ethical considerations around the use of data in training the models, along with privacy considerations, might also be worthy of future examination.

References

[1]. Amshuman Singh (August 2021)-"Machine Learning Approach to Crime Prediction and Identification of Hotspots".

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

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 04 April 2025
Page No: 2007-2015
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0293

[3]. Md Manowarul Islam, Md Ashraf Uddin, Linta Islam (2020)- "Cyberbullying Detection on Social Networks Using Machine Learning Approaches".
[4]. Bandeh Ali Talpur, Declan O'Sullivan (October 2020)- "Cyberbullying severity detection: A machine learning approach".
[5]. Department of Translation, Interpreting, and Communication - Faculty of Arts and Philosophy, Ghent University, Ghent, Belgium, (October 2018), "Automatic detection of cyberbullying in a social media text".
[6]. John Hani, Mohamed Nashaat, Mostafa Ahmed, Zeyad Emad, Eslam Amer, Ammar Mohammed, (2019), "Social Media Cyberbullying Detection using Machine Learning".
[7]. Nureni Ayofe Azeez, Sunday O. Idiakose, Chinazo Juliet Onyema and Charles Van Der Vyver, (June 2021), "Cyberbullying Detection in Social Networks: Artificial Intelligence Approach".
[8]. Afrah Almansoori, Mohammed Alshamsi, Sherief Abdallah, and Said A. Salloum, (2021), "Analysis of Cybercrime on Social Media Platforms and Its Challenges".