Detecting LGBTQ+ Instances of Cyberbullying

Muhammad Arslan[1], Manuel Sandoval Madrigal[1], Mohammed Abuhamad[1], Deborah L. Hall[2], and Yasin N. Silva[1]

[1] Loyola University Chicago, Chicago IL 60626, USA
{marslan,msandovalmadrigal,mabuhamad,ysilva1}@luc.edu
[2] Arizona State University, Glendale, AZ 85306
d.hall@asu.edu

Abstract. Social media continues to have an impact on the trajectory of humanity. However, its introduction has also weaponized keyboards, allowing the abusive language normally reserved for in-person bullying to jump onto the screen, i.e., cyberbullying. Cyberbullying poses a significant threat to adolescents globally, affecting the mental health and well-being of many. A group that is particularly at risk is the LGBTQ+ community, as researchers have uncovered a strong correlation between identifying as LGBTQ+ and suffering from greater online harassment. Therefore, it is critical to develop machine learning models that can accurately discern cyberbullying incidents as they happen to LGBTQ+ members. The aim of this study is to compare the efficacy of several transformer models in identifying cyberbullying targeting LGBTQ+ individuals. We seek to determine the relative merits and demerits of these existing methods in addressing complex and subtle kinds of cyberbullying by assessing their effectiveness with real social media data.

Keywords: Cyberbullying · LGBTQ+, · Social Media · LLMs

1 Introduction

Cyberbullying continues to be a pressing issue that affects a considerable proportion of adolescents across the world. It is estimated that nearly 10% to 20% of adolescences have experienced cyberbullying at some time, and if left unaddressed, cyberbullying can contribute to significantly increased risk for suicide among adolescents [1]. That is, while the exact relationship between bullying and suicidality (suicidal ideation or suicide attempts) is complex, there is evidence to suggest that exposure to peer bullying and peer aggression are risk factors [2]. As with many of the technological advancements of our day, bullying has also evolved into cyberbullying, the use of technology/electronic media as a means to bully or harass. One way to address cyberbullying is through the development of machine learning-based moderation tools that readily scale for the task of cyberbullying detection. While prior research in cyberbullying detection has focused primarily on the issue of general cyberbullying, limited work has been done towards developing LGBTQ+ sensitive models. Specifically, most cyberbullying detection models approach cyberbullying as one-size-fits-all, while the reality is

that the type of harassment experienced by LGBTQ+ members varies significantly from that of the general public. Even more concerning is that sexual and gender identity minorities are significantly more likely to be victims of cyberbullying and online harassment compared to heterosexual and cisgender individuals [3]. This strife is then compounded as LGBTQ+ members are more severely impacted by these negative interactions due to well-documented disparities in mental health and social support and the chronic experiences of minority stress [4]. With cyberbullying corresponding with an increase in suicidality amongst the general public, we see even more pressing need for cyberbullying models that support the LGBTQ+ community.

In this paper, we describe the implementation and performance evaluation of several large language-based models that aim to identify cyberbullying posts that target LGBTQ+ individuals. We seek to determine the relative merits and demerits of these existing methods in addressing complex and subtle kinds of cyberbullying by assessing their effectiveness with an Instagram dataset.

2 Related Work

Cyberbullying Detection via Machine Learning. The domain of general cyberbullying detection has received considerable empirical attention, in terms of efforts to adapt the latest methods from the machine learning literature and to propose models that take advantage of features specific to each social media platform. Early contributions to cyberbullying detection focused on applying off-the-shelf solutions, e.g., SVM, Naïve Bayes, and Logistic Regression, to standard binary classification (bullying versus non-bullying) [5]. Well-established deep learning architectures, such as CNN, LSTM, BiLSTM, and BiLSTM with attention, were studied and applied by Dadvar and Eckert to a cyberbullying-labeled YouTube dataset that included 53k posts and 4k users [6]. Cheng et al. proposed the inclusion of network-related content such as user profile information, likes, and follows to identify cyberbullying [7]. In other studies by Cheng et al. [8,9], they studied how the temporal dynamics of comment arrival can be modeled as part of a hierarchical attention network. Some researchers have also proposed multi-modal models that incorporate video, images, and time-related components into the architectures of their models [10].

Cyberbullying Risk among LGBTQ+ Users. A robust finding within social science research is the disproportionately high rates at which LGBTQ+ individuals are targets of bullying, harrassment, and discrimination—including cyberbullying (e.g., [3]). For example, in a national sample of U.S. teens, more than half of those identifying as LGBTQ+ (52%) had been a victim of cyberbullying compared to 35% of Non-LGBTQ+ teens [3]. The importance of efforts to help identify and quickly respond to cyberbullying targeting LGBTQ+ users is underscored by mental health disparities that are distinct from and also exacerbated by cyberbullying experiences. That is, relative to their Non-LGBTQ+ peers, LGBTQ+ individuals experience poorer mental health outcomes [4], due

in part to chronic experiences of minority stress and broader socioeconomic and health disparities ([11]), and are more likely to suffer negative mental health consequences, including depression, as a result of cyberbullying [12].

Toward LGBTQ+ Cyberbullying Detection. Most of the previous work on cyberbullying detection has focused on the development of generic models. However, recent work highlighted potential biases against certain users that can be introduced by available datasets and previous models [13]. For instance, this work reported that 68.4% of sessions containing the word "gay" were labeled bullying in a commonly used Instagram dataset [14]. Moreover, Tangila et al. highlighted the importance of tailored online safety tools after evaluating an Instagram dataset and finding that LGBTQ+ teens experienced significantly more high-risk online interactions than their heterosexual peers [15]. More recently, some initial work aimed at automatically detecting homophobia and transphobia in YouTube comments [16] reported the results of a shared task conducted as part of a workshop in homophobia and transphobia detection (LTEDI-ACL 2022). The top-2 results for English obtained F1 Macro scores were 57% and 49% (based on the work by Maimaitituoheti et al. [17] and Sammaan et al. [15], respectively), highlighting the need for additional work in this area.

3 Methods

Problem Definition. Let $C = \{p_1, p_2, \ldots, p_n\}$ be a corpus of n samples, where $p_i = \{w_1, w_2, \ldots, w_l\}$ is the l-length tokenized representation of a given comment. Let $Y = \{y_1, y_2, \ldots, y_n\}$ be the associated labels for samples in C where each $p_i \rightarrow y_i$. For any label $y_i \in \{0, 1\}$, we have $y_i = 1$ when a comment contains LGBTQ+ related cyberbullying and $y_i = 0$ when a comment does not. The goal is to train a LGBTQ+-sensitive classifier f such that $f(p) \rightarrow y$.

Dataset. The dataset used in this research was obtained from the study "A Labeled Dataset for Investigating Cyberbullying Content Patterns in Instagram" [18]. The dataset contains 1,083 Instagram comments related to cyberbullying, with 217 comments specifically targeting the LGBTQ+ community. The dataset was annotated by a diverse team of annotators with expertise in psychology and computer science, providing detailed labels at both the session and comment levels. These labels capture essential aspects of cyberbullying, such as content type, purpose, directionality, and co-occurrence with other phenomena. The choice of this dataset aligns with our research objectives and enables an effective analysis of cyberbullying patterns on real-world social media platforms. An example of the comments that can be found in this dataset is presented in Table 1

Model Selection. We consider three pre-trained language models, namely Roberta [19], BERT [20], and GPT-2 [21], to classify cyberbullying comments as either LGBTQ-related or Non-LGBTQ-related. For each model, we experimented with various configurations to evaluate their performance and robustness. These configurations include the original dataset without oversampling, oversampling using SMOTE (Synthetic Minority Over-sampling Technique) [22], and oversampling using ADASYN (Adaptive Synthetic Sampling) [23].

LGBTQ+ Related Comments	**Non-LGBTQ+ Related Comments**
Hey everyone this faggot @**username1** is following 666 people.	@**username2** lmao just because ur an elounor shipper doesn't mean you have to be a bitch lol shut up
He must be illuminati right? Do your research before you input your opinion. You sound ignorant.	And one more thing @**username3** whoooooo cares about your st**** boyfriend.....whell i think NONE OF USS exept u :)
Ur gay dawg like really	Fuck you! And btw, it was the X factor nimrod. She can sing better than you will in your life time. So stfu. @**username4**
PUNK ASS DADDY FAG	Ur a white piece of trash
Gay people disgust me and I hate them. They scared me! That's all I'm gunna say!	Shove off baby ugly @**username5**

Table 1. A sample of both LGBTQ+ and non-LGBTQ+ cyberbullying comments. Usernames were anonymized.

Model Training and Evaluation. For this study we considered the models, Roberta, BERT, and GPT-2, which were sourced from the Transformers library. We studied each model in an identical testing pipeline which includes preprocessing and tokenization using the model specific tokenizers, training the models to the desired number of epochs, and evaluating model performance based Accuracy, Precision, Recall, F1 score, and AUROC. For all the models, we utilized both SMOTE and ADASYN oversampling techniques with 5-fold cross-validation to address the potential class imbalance in the dataset and ensure the models perform consistently regardless of the training-validation split. The aforementioned oversampling techniques generate additional samples of the minority class to balance the class distribution and present the model with more opportunities to learn the underrepresented class. At the end of the training pipeline, each model has been tested across three experiment configurations, covering all possible combinations of the proposed oversampling techniques (original, SMOTE, and ADASYN).

Model Fine-tuning. We used a fine-tuning approach to enhance the performance of pre-trained language models for classifying LGBTQ+ cyberbullying comments. To use RoBERTa and BERT, these models were adapted with sequence classification heads. The models were then trained using the binary cross-entropy loss function and optimized with the AdamW optimizer. The learning rate was set to $5e-5$ and the models were trained for 3 epochs with a batch size of 16. Adapting GPT-2 for classification required additional steps since it was originally designed for text generation. We utilized the GPT2ForSequenceClassification which adds a linear layer on top of the pooled output of GPT-2 enabling it to perform sequence classification. To handle variable-length sequences we used the

EOS token as the padding token and set the maximum sequence length to 128 tokens.

Evaluation Metrics and Visualization. To assess the performance of the models, we used a range of evaluation metrics including accuracy, precision, recall, F1 score, and Area Under the Receiver Operating Characteristic curve (AUROC). These metrics provide a holistic view of the models' performance, considering different aspects such as correct classifications, false positives, and false negatives. In addition to quantitative metrics, we generated confusion matrices and visualization plots for each fold of the cross-validation. The performance metrics presented in the Experimental Results are the result of taking the average score across the 5 folds.

4 Experimental Results

The experimental results of the three pre-trained language models; Roberta, BERT and GPT-2, for classifying cyberbullying comments as either LGBTQ-related or Non-LGBTQ-related are presented in Table 2.

MODEL	Oversampling	Metrics				
		Accuracy	Precision	Recall	F1	AUROC
RoBERTa	Original	**0.9456**	0.8712	**0.6372**	**0.733**	0.8592
	SMOTE	0.938	**0.8888**	0.5664	0.6826	0.8376
	ADASYN	0.9412	0.8658	0.6134	0.7044	**0.8926**
BERT	Original	0.8922	0.4734	0.1092	0.172	0.7164
	SMOTE	0.941	0.8792	0.5912	0.6942	0.8626
	ADASYN	0.9298	0.8688	0.4786	0.611	0.871
GPT-2	Original	0.9344	0.8556	0.5338	0.6506	0.8776
	SMOTE	0.9032	0.7684	0.213	0.3218	0.7524
	ADASYN	0.9114	0.8162	0.3332	0.4468	0.8078

Table 2. Comparison of Different Models Across Various Metrics

The RoBERTa model outperforms BERT and GPT-2 across most configurations, achieving the highest accuracy (0.9456) and F1 (0.733) scores without oversampling. The overall best AUROC score (0.8926) was also obtained with RoBERTa but using ADASYN oversampling. These results indicate RoBERTa's robustness in classifying LGBTQ-related cyberbullying comments. However, there are notable challenges and limitations in the models' performance, particularly in identifying cyberbullying instances, which we discuss below.

Despite the overall strong performance, the models, including RoBERTa, exhibit difficulties in accurately identifying cyberbullying instances, especially those targeted at LGBTQ+ individuals. Table 3 provides a per-class breakdown of precision, recall, F1 score, and AUROC for RoBERTa with different oversampling techniques.

Model	Oversampling	Class	Metrics			
			Precision	Recall	F1	AUROC
RoBERTa	Original	Non-LGBTQ+	0.952	0.986	0.968	0.8592
		LGBTQ+	0.8712	0.6372	0.733	
	SMOTE	Non-LGBTQ+	0.938	0.994	0.962	0.8376
		LGBTQ+	0.8888	0.5664	0.6826	
	ADASYN	Non-LGBTQ+	0.95	0.982	0.968	0.8926
		LGBTQ+	0.8658	0.6134	0.7044	

Table 3. Per-class Comparison of Roberta Models with Different Oversampling Techniques Across Various Metrics

The per-class analysis highlights that although RoBERTa demonstrates strong performance in accurately identifying Non-LGBTQ+ bullying comments, as evidenced by the high precision and recall scores, it faces challenges when it comes to detecting LGBTQ+ bullying comments. This is reflected in the consistently lower recall scores for the bullying class across all experimental configurations. The model's difficulty in recognizing subtle and context-dependent instances of bullying leads to an increased number of false negatives, where LGBTQ+ bullying comments are incorrectly classified as Non-LGBTQ+ bullying. This limitation underscores the need for further research and development to enhance the model's ability to capture the nuances and complexities of LGBTQ+-related cyberbullying.

To further illustrate the models' performance, confusion matrices for each configuration are presented in Figure 1. This figure includes the confusion matrices for each model and oversampling technique combination. The confusion matrices provide a comprehensive overview of the models' performance by displaying the actual and predicted values for both the Non-LGBTQ+ bullying (0) and LGBTQ+ bullying (1) classes.

Across all models and oversampling techniques, the true negative values (top-left) are consistently higher than the true positive values (bottom-right) indicating that the models perform better in identifying Non-LGBTQ+ bullying comments compared to LGBTQ+ bullying comments. The visualization also highlights the impact of oversampling techniques on the models' performance. For instance, SMOTE and ADASYN oversampling generally lead to a higher number of true positives compared to no oversampling, suggesting an improvement in the models' ability to detect LGBTQ+ bullying instances.

However, the confusion matrices also reveal potential areas for improvement, particularly in reducing the number of false negatives (LGBTQ+ bullying comments incorrectly classified as Non-LGBTQ+ bullying). This is evident from the relatively high values in the bottom-left cells of the matrices. Addressing this issue is crucial for developing more effective models that can accurately identify and mitigate cyberbullying targeting the LGBTQ+ community.

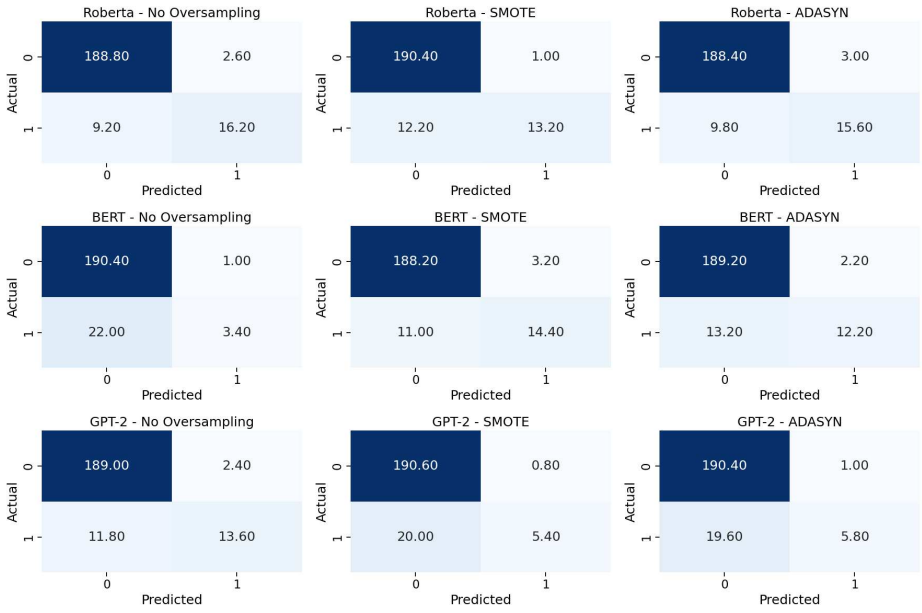

Fig. 1. Above is a set of confusion matrices, one for each model and oversampling technique combination.

5 Discussion

The experimental results presented in this study highlight the potential and limitations of state-of-the-art transformer models (RoBERTa, BERT and GPT-2) for detecting LGBTQ+-related cyberbullying comments. While the models demonstrate overall effectiveness in distinguishing between Non-LGBTQ+ bullying and LGBTQ+ bullying comments, several technical aspects warrant further discussion.

Performance Evaluation and Challenges. RoBERTa consistently outperforms BERT and GPT-2 across various metrics, particularly in terms of accu-

RoBERTA Mislabeled Comments	**Outcome**
No more calling me white trash or fag @p3bble	FP
@bacon I agree and @icrazyfan190 your an ass, gay people are still people and everyone should learn to respect them because we are all equal	FP
@greenlady your what I like to call a top class bell end. You are a dyke!	FN
U like dick #thatsurfuckinproblem	FN

Table 4. Samples of both False Positive (FP) and False Negative (FN) comments from RoBERTa.

racy and F1 scores. However, the confusion matrices reveal a disparity between true negatives (Non-LGBTQ+ bullying comments) and true positives (LGBTQ+ bullying comments), highlighting the models' tendency to favor Non-LGBTQ+ bullying classifications, leading to higher false negatives for LGBTQ-related Cyberbullying. In Table 4, we see can see two examples of misclassified false negatives comments. The first false negative might indicate that the model did not recognize the homophobic slur 'dyke', which is likely a symptom of insufficient training data. In the second example, the model likely struggled with the lack of context in the comment to confidently label it as LGBTQ+ bullying. The persistent challenge in accurately identifying LGBTQ+-related cyberbullying comments lies in the nuanced and context-dependent nature of such interactions. Cyberbullying targeting LGBTQ+ individuals often involves implicit language, sarcasm, and coded expressions, e.g., the false positives and false negative examples present in Table 4, that are difficult for models to decipher, resulting in lower precision and recall rates for the bullying class.

Impact of Oversampling Techniques. The integration of oversampling techniques such as SMOTE and ADASYN shows a noticeable impact on the models' ability to detect bullying comments by providing more balanced training data. However, the issue of false negatives remains prevalent, suggesting that oversampling alone is insufficient to address the inherent complexities of LGBTQ+-related cyberbullying detection.

Future Work. Several techniques could be explored to further improve the performance of LGBTQ+ cyberbullying detection. These include: (1) integrating data features and model mechanisms that can capture deeper contextual and semantic nuances, (2) integrating multi-modal data including images, videos, and social network metrics (e.g., likes and shares) to enrich contextual data, (3) creating datasets that reflect a wider range of bullying scenarios, particularly those targeting LGBTQ+ individuals, and (4) exploring models that account for the sequence and timing of interactions to better capture intrinsic cyberbullying properties.

Limitations of the Current Study. The study's limitations include the dataset size and diversity and the single-platform focus on Instagram data. Addressing these limitations in future research is essential for developing more generalizable and effective cyberbullying detection systems.

6 Conclusion

This study evaluated RoBERTa, BERT, and GPT-2 for detecting LGBTQ+ cyberbullying on social media. RoBERTa outperformed BERT and GPT-2, but all models struggled to some extent with nuanced, context-dependent bullying instances. Oversampling techniques like SMOTE and ADASYN improved detection but did not fully resolve false negatives. Future work could focus on leveraging larger, more diverse datasets, incorporating multi-modal data, and developing fairness-aware training methodologies to enhance model effectiveness and create safer online spaces for the LGBTQ+ community.

Acknowledgments. This work was supported by NSF Awards #2227488 and #1719722 and a Google Award for Inclusion Research.

References

1. G. Gini and D. L. Espelage, "Peer Victimization, Cyberbullying, and Suicide Risk in Children and Adolescents," *JAMA*, vol. 312, no. 5, pp. 545–546, 08 2014. [Online]. Available: https://doi.org/10.1001/jama.2014.3212
2. M. K. Holt, A. M. Vivolo-Kantor, J. R. Polanin, K. M. Holland, S. DeGue, J. L. Matjasko, M. Wolfe, and G. Reid, "Bullying and Suicidal Ideation and Behaviors: A Meta-Analysis," *Pediatrics*, vol. 135, no. 2, pp. e496–e509, 02 2015. [Online]. Available: https://doi.org/10.1542/peds.2014-1864
3. C. R. Centre, "Bullying, cyberbullying, and lgbtq students," 2020, may 28, 2024. [Online]. Available: https://www.humanitarianlibrary.org/resource/bullying-cyberbullying-and-lgbtq-students
4. M. Plöderl and P. Tremblay, "Mental health of sexual minorities. a systematic review," *International Review of Psychiatry*, vol. 27, no. 5, pp. 367–385, 2015, pMID: 26552495. [Online]. Available: https://doi.org/10.3109/09540261.2015.1083949
5. M. Dadvar, F. de Jong, R. Ordelman, and R. Trieschnigg, "Improved cyberbullying detection using gender information," in *Proceedings of the Twelfth Dutch-Belgian Information Retrieval Workshop (DIR 2012)*. Belgium: Ghent University, 2012, pp. 23–25.
6. M. Dadvar and K. Eckert, "Cyberbullying detection in social networks using deep learning based models; a reproducibility study," *arXiv preprint arXiv:1812.08046*, 2018.
7. L. Cheng, J. Li, Y. Silva, D. Hall, and H. Liu, "PI-bully: Personalized cyberbullying detection with peer influence," *Electronic proceedings of IJCAI 2019*, pp. 5829–5835, 2019.
8. L. Cheng, R. Guo, Y. Silva, D. Hall, and H. Liu, "Hierarchical attention networks for cyberbullying detection on the instagram social network," in *Proceedings of the 2019 SIAM International Conference on Data Mining (SDM)*. Society for Industrial and Applied Mathematics, 2019, pp. 235–243.
9. L. Cheng, R. Guo, Y. N. Silva, D. Hall, and H. Liu, "Modeling temporal patterns of cyberbullying detection with hierarchical attention networks," *ACM/IMS Trans. Data Sci.*, vol. 2, no. 2, 2021.
10. V. K. Singh, S. Ghosh, and C. Jose, "Toward multimodal cyberbullying detection," in *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. Association for Computing Machinery, 2017, pp. 2090–2099.
11. T. Dürrbaum and F. A. Sattler, "Minority stress and mental health in lesbian, gay male, and bisexual youths: A meta-analysis," *Journal of LGBT Youth*, vol. 17, no. 3, pp. 298–314, 2020. [Online]. Available: https://doi.org/10.1080/19361653.2019.1586615
12. C. Duarte, S. K. Pittman, M. M. Thorsen, R. M. Cunningham, and M. L. Ranney, "Correlation of minority status, cyberbullying, and mental health: A cross-sectional study of 1031 adolescents," *Journal of Child & Adolescent Trauma*, vol. 11, no. 1, pp. 39–48, 2018. [Online]. Available: https://doi.org/10.1007/s40653-018-0201-4

13. L. Cheng, A. Mosallanezhad, Y. Silva, D. Hall, and H. Liu, "Mitigating bias in session-based cyberbullying detection: A non-compromising approach," in *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, C. Zong, F. Xia, W. Li, and R. Navigli, Eds. Online: Association for Computational Linguistics, Aug. 2021, pp. 2158–2168. [Online]. Available: https://aclanthology.org/2021.acl-long.168
14. H. Hosseinmardi, S. A. Mattson, R. I. Rafiq, R. Han, Q. Lv, and S. Mishra, "Detection of cyberbullying incidents on the instagram social network," 2015.
15. T. I. Tanni, M. Akter, J. Anderson, M. J. Amon, and P. J. Wisniewski, "Examining the unique online risk experiences and mental health outcomes of lgbtq+ versus heterosexual youth," in *Proceedings of the CHI Conference on Human Factors in Computing Systems*, ser. CHI '24. ACM, May 2024. [Online]. Available: http://dx.doi.org/10.1145/3613904.3642509
16. B. R. Chakravarthi, R. Priyadharshini, R. Ponnusamy, P. K. Kumaresan, K. Sampath, D. Thenmozhi, S. Thangasamy, R. Nallathambi, and J. P. McCrae, "Dataset for identification of homophobia and transophobia in multilingual youtube comments," 2021.
17. A. Maimaitituoheti, "ABLIMET @LT-EDI-ACL2022: A roberta based approach for homophobia/transphobia detection in social media," in *Proceedings of the Second Workshop on Language Technology for Equality, Diversity and Inclusion*, B. R. Chakravarthi, B. Bharathi, J. P. McCrae, M. Zarrouk, K. Bali, and P. Buitelaar, Eds. Dublin, Ireland: Association for Computational Linguistics, May 2022, pp. 155–160. [Online]. Available: https://aclanthology.org/2022.ltedi-1.19
18. M. Hamlett, G. Powell, Y. N. Silva, and D. Hall, "A labeled dataset for investigating cyberbullying content patterns in instagram," *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 16, no. 1, pp. 1251–1258, May 2022. [Online]. Available: https://ojs.aaai.org/index.php/ICWSM/article/view/19376
19. Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, "Roberta: A robustly optimized bert pretraining approach," 2019.
20. J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," 2019.
21. A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, and I. Sutskever, "Language models are unsupervised multitask learners," 2019.
22. N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "Smote: Synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, vol. 16, p. 321–357, Jun. 2002. [Online]. Available: http://dx.doi.org/10.1613/jair.953
23. H. He, Y. Bai, E. A. Garcia, and S. Li, "Adasyn: Adaptive synthetic sampling approach for imbalanced learning," in *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, 2008, pp. 1322–1328.