Paper Title: Cyberbullying Detection Using Deep Neural Network from Social Media Comments in Bangla Language.

Paper Link: https://arxiv.org/ftp/arxiv/papers/2106/2106.04506.pdf

1. Summary

1.1 Motivation/ Purpose/ Aim

The prevalence of social media usage in the Bangla-speaking world has led to a significant increase in cyberbullying. Using a deep neural network model, the authors are motivated to develop a cyberbullying detection system to protect Bangla-speaking social media users. Because the authors hypothesize that a deep neural network model can be trained to accurately detect cyberbullying in Bangla social media comments, the purpose of this paper is to propose a deep neural network model for cyberbullying detection in Bangla social media comments. As stated previously, the purpose of this paper is to construct the deep neural network model mentioned by the authors, evaluate its performance on a real-world dataset, and then make the dataset and code publicly accessible to facilitate further research.

1.2 Contribution

1. They suggested using a deep neural network model to find abuse in comments on Bangla social media sites. It is a CNN-LSTM model, which means it takes the best parts of both CNNs and LSTMs.
2. They tested how well the suggested model worked on a real-world dataset of 44,001 comments in Bangla from public Facebook pages. The model was right 87.91% of the time when it came to binary classification (Bully vs. Non-bully) and 85% of the time when it came to multi-class classification (Non-bully, Sexual, Threat, Troll, and Religious).
3. They shared the dataset and code with the public so that more study could be done on finding cyberbullying in Bangla social media comments.

1.3 Methodology

Dataset collection and labeling were performed initially. The authors then extracted features from the comments and represented them in a format that the neural network model could understand using a variety of techniques. Among these methods were named entity recognition, part-of-speech labeling, and word embedding. In addition, the authors devised a CNN-LSTM hybrid model with the purpose of detecting cyberbullying. On the gathered dataset, the Adam optimizer and the cross-entropy loss function were utilized to train the model. The CNN is employed to derive features from the comments, while the LSTM is utilized to discover the text's long-range dependencies. In conclusion, the researchers assessed the effectiveness of the suggested model by applying the subsequent metrics to the gathered data set: F1 score, accuracy, precision, and recall. In addition, the authors conducted experiments utilizing various architectures of neural networks, including RoBERTa and Bidirectional Encoder Representations

from Transformers (BERT). On the collected dataset, they discovered that the proposed hybrid CNN-LSTM model outperformed the other architectures.

1.4 Conclusion

The authors end the research by claiming that the suggested deep neural network model is a viable solution for cyberbullying identification in Bangla social media comments. The model performed well on the obtained data, suggesting it might be utilized to aid in the creation of a cyberbullying detection system tailored to the needs of Bangla-speaking social media users.

2 Limitations

2.1 First Limitation

The model underwent evaluation exclusively on a single dataset sourced from publicly accessible Facebook profiles. In a more realistic environment, it is essential to evaluate the model's performance using additional datasets, including private social media posts and conversations.

2.2 Second Limitation

The model doesn't look at any background information, like how the people talking to each other are related or how they've interacted with each other before. Adding relevant information to the model could make it more effective in identifying cyberbullying.

3 Synthesis/ Future work

1. Evaluating the proposed model on other datasets to assess its generalizability.
2. Enhancing the system by integrating additional supplementary characteristics, such as user profiles and social network context, into the model.
3. Implementing the suggested model for cyberbullying detection in different languages, including Hindi, English, and Spanish
4. Implementing the suggested framework within social media platforms to assist them in the identification and removal of cyberbullying material.