Paper title:
Abusive Bangla comments detection on Facebook using transformer-based deep learning models.

Paper link:
https://link.springer.com/article/10.1007/s13278-021-00852-x

1 Summary:

1.1 Motivation:

The paper aims to detect abusive comments in Bangla language on Facebook using transformer-based deep learning models. The motivation behind this research is to identify and filter out abusive comments at the primitive stage of social media's affixing, as online content with extreme toxicity can have negative effects on individuals and society.

1.2 Contribution:

The paper contributes to the field of natural language processing by proposing a model that can classify abusive comments in Bangla language with high accuracy. The authors also provide a dataset of Bangla comments labeled as abusive or non-abusive, which can be used for further research in this area.

1.3 Methodology:

The proposed framework for detecting abusive comments in Bangla language on Facebook using transformer-based deep learning models involves preprocessing the comment texts from the dataset before training the model. The data preprocessing involves several steps, including removing stop words, stemming, and tokenization. The comments are then converted into numerical vectors using word embeddings. The resulting vectors are then fed into the transformer-based deep learning models for training and classification. Detecting abusive comments in Bangla language on Facebook using transformer-based deep learning models involves using both BERT and ELECTRA models for training and classification. The model is trained using preprocessed comments and transformer-based learning to classify abusive comments. Finally, the model is fine-tuned using different values of hyperparameters to predict classes more accurately.

1.4 Conclusion:

The proposed model achieved high accuracy in detecting abusive comments in Bangla language on Facebook. The authors suggest that this model can be used to filter out abusive comments at the primitive stage of social media's affixing, which can have a positive impact on individuals and society.

2 Limitations:

2.1 First Limitation/Critique:

The dataset used in this research is limited to Bangla comments on Facebook, which may not be representative of all Bangla comments on social media. The model's performance may vary when applied to other social media platforms or languages.

2.2 Second Limitation/Critique:
The authors did not consider the context of the comments while labeling them as abusive or non-abusive. This may lead to misclassification of comments that are not necessarily abusive but may appear so out of context.

3 Synthesis:
The proposed model can be applied to other languages and social media platforms to detect abusive comments. This can have a positive impact on individuals and society by filtering out toxic content at the primitive stage of social media's affixing. Future research can focus on improving the model's performance by considering the context of comments and expanding the dataset to include comments from other social media platforms and languages.