Detecting Cyberbullying on Twitter Using Natural language processing Techniques and Deep learning models

Zandile Queeneth Shabangu (154632)
Supervisor: Dr. E. Mbunge
Department of Computer Science and, Engineering University of Eswatini

Abstract

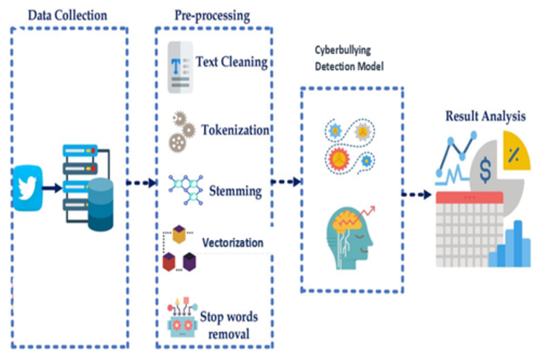
Social media platforms present tremendous opportunities for people to share their sentiments, emotions, and opinions in the virtual space. However, some users utilize social media platforms and engage in malicious behaviour such as cyberbullying or hate speech. Cyberbullying is anonymous and quite difficult to detect in real-time. This has devastating psychological effects which consequently lead to anxiety, depression, and even an increase in suicide cases. Therefore, there is a need to develop cyberbullying intelligent detection systems to alleviate this problem. In this study, we applied deep learning models to detect cyberbullying using real-time Twitter data. Data were extracted using the most prominent BlackLivesMatter, a social movement associated with hate speech, cyberbullying, misinformation, and racial discrimination.

1. Introduction

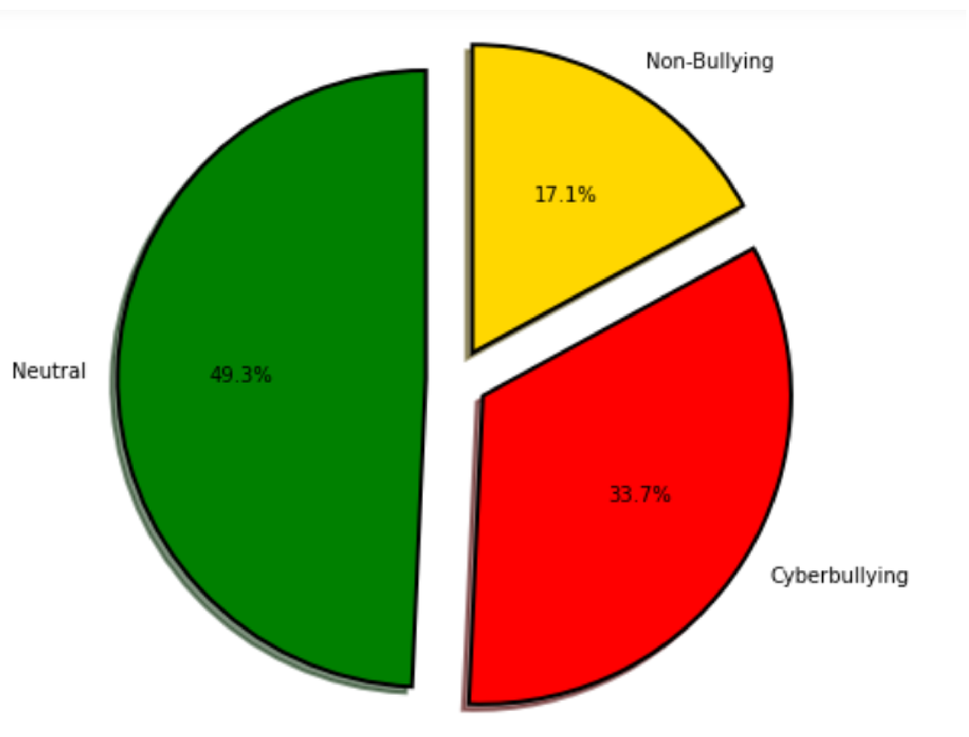
- Natural language processing (NLP) is the ability of a model/ algorithm to understand human language as it is spoken and written - referred to as natural language.
- It is a component of artificial intelligence (AI).
- Deep Learning is a subset of AI that supports hierarchical representation of data and abstractions (Shinde, Shah, 2018).
- Sentiment analysis, also known as opinion mining, or emotion AI, focuses on analyzing online data from various social media platforms to determine the emotional tone (Ligthart et al., 2021).
- Cyberbullying- is used to describe many different kinds of online abuse such as harassment, hate speech, insults, and racism perpetrated by malicious social media users (Dadvar, Eckert, 2018).
- Several authors applied Natural Language Processing and deep learning techniques to detect cyberbullying. Deep learning models have been substantially used in various social media applications, including sentiment analysis and natural language processing tasks, for different purposes, including sentiment classification. This is because classic machine learning techniques have a limited capability to efficiently analyze such large amounts of data and produce precise results; they are thus supported by deep learning models to achieve higher accuracy.

2. Method

- Data preprocessing and cleaning were done by applying tokenization, stemming, and vectorization.
- After cleaning data, the study applied Bi-directional Long Short Term Memory (BLSTM), deep neural networks (DNN), and Gated Recurrent Units (GRU) to detect cyberbullying.
- Accuracy, precision, recall and F1-measure were used to assess the performance of deep learning-based cyberbullying detection models.



3. Sentiments Results



4. Key Results

- BLSTM achieved an accuracy of 85.71 percent, precision of 85.71 percent, recall of 85.71 percent, and F1-measure of 85.71 percent.
- Deep neural networks achieved an accuracy of 66.03, precision of 81.82 percent, recall of 54.29 percent, and F1-measure of 65.27 percent.
- GRU achieved an accuracy of 66.03 percent, precision of 81.82 percent, recall of 54.29 percent, and F1-measure of 65.27 percent.

5. Classification Report

Model	Accuracy	Precision	Recall	F1Measure
BLSTM	85.71	85.71	85.71	85.71
DNN	66.03	81.82	54.29	65.27
GRU	66.03	81.82	54.29	65.27

6. Conclusion and Future Work

- The study findings may inspire policymakers and law enforcement agencies to leverage deep learning models to build deployable intelligent systems to detect cyberbullying.
- These models could be embedded in social media platforms to detect cyberbullying in real-time.
- Future work may focus on developing models for detecting cyberbullying in hidden sarcasm images and posts written in vernacular languages.

References

1. Shinde, P. P., Shah, S. (2018). A Review of Machine Learning and Deep Learning Applications. Proceedings - 2018 4th International Conference on Computing, Communication Control and Automation, ICCUBEA 2018.
2. Park, H. J., Francisco, S. C., Pang, M. R., Peng, L., Chi, G. (2021). Exposure to anti-Black Lives Matter movement and obesity of the Black population. Social Science Medicine, 11426.
3. 5Crowder, C. (2021). When BlackLivesMatter at the Women's March: a study of the emotional influence of racial appeals on Instagram.