Analyzing Home Violence Incidents using Social Media: A Case Study on Twitter

Adeeba S.[1*], Kumara B.T.G.S.[1], and Banujan K.[1]

[1]Department of Computing and Information Systems, Faculty of Computing, Sabaragamuwa University of Sri Lanka, Sri Lanka
[*]adeeadee2345@gmail.com

The outbreak of COVID-19 has set off a worldwide well-being disaster that affects how we view the world and direct our everyday lives. The impact has given both positive and negative results. One of the negative results of the COVID-19 pandemic is Home Violence (HV). HV encompasses many misuses, including physical abuse, sexual abuse, emotional abuse, and controlling behavior in a close relationship. People are becoming more reliant on social media platforms like Twitter, Instagram, Facebook, YouTube, etc. Twitter has recently emerged as an excellent resource for studying COVID-19 user-generated material and behaviors in real-time. Analyzing HV-related posts on social media is beneficial in gauging public sentiment toward sensitive problems, public expression of feelings, and resource sharing regarding the otherwise personal experience of HV. Our research proposed a method to analyze the HV incidents using social media during the COVID-19 pandemic. More than 20,000 Tweets were retrieved between 2020 April to 2021 July using Twitter API. Data pre-processing and word embedding were done, respectively. Then, to construct the model, the data set was split into training and testing datasets to detect HV-related Tweets; a deep learning model, LSTM, with different word embedding techniques, was used in this research (TF- IDF+LSTM, BOW+LSTM, Word2Vec+LSTM, GloVe+LSTM, and BERT+LSTM). After, HV-related Tweets are classified into three main topics: HV incident, HV awareness, and HV shelter with the help of LSTM with GloVe embedding. Finally, 5W proposed a model introduced to describe the HV incident Tweets and It's including 'What', 'When', 'Where', 'Who', and 'Why' elements. With an accuracy of 89.56%, the BERT+LSTM model surpassed the other implemented models. The proposed GloVe+LSTM achieved an accuracy of 98.35% to classify the HV Tweets into three main categories. HV incidents Tweets reveal that the proposed 5W model performs well in describing the HV incidents.

Keywords: *Social media, Twitter, Domestic violence, Deep learning, Word embedding*