|  |  |  |  |
| --- | --- | --- | --- |
| Title | How can we manage offensive text in social media-a text classification approach using LSTM-BOOST | | |
| Author(s) Name | Md Anwar Hussen Wadud, Muhammad Mohsin Kabir, MF Mridha, M Ameer Ali, Md Abdul Hamid, Muhammad Mostafa Monowar | | |
| Contact Email(s) | firoz.mridha@aiub.edu | | |
| Published Journal Name | International Journal of Information Management Data Insights | | |
| Type of Publication | Journal | | |
| Volume | 2 | Issue | 2 |
| Publisher | Elsevier | | |
| Publication Date | 2022/11/1 | | |
| ISSN | 0346-251X | | |
| DOI | 10.1016/j.jjimei.2022.100095 | | |
| URL | https://www.sciencedirect.com/science/article/pii/S2667096822000386 | | |
| Other Related Info. |  | | |
|  | | | |

|  |  |
| --- | --- |
| Abstract |  |
| Recently, offensive content has become increasingly popular for harassing and criticizing people on numerous social media platforms. This paper proposes an offensive text classification algorithm named LSTM-BOOST employing Long Short-Term Memory(LSTM) model with ensemble learning to recognize offensive Bengali texts in various social media platforms. The proposed LSTM-BOOST model uses the modified AdaBoost algorithm employing principal component analysis(PCA) along with LSTM networks. In the LSTM-Boost model, the dataset is divided into three categories, and PCA and LSTM networks are applied to each part of the dataset to obtain the most significant variance and reduce the weighted error of the weak hypothesis of the model. Furthermore, different classifiers are used for baseline experiment and the model is evaluated on various word embedding vector methods. Our investigation found that the LSTM-BOOST algorithms outperform most of the baseline architecture, leading F1-score of 92.61% on the Bengali offensive text from Social Platforms(BHSSP) dataset. | |