|  |  |  |  |
| --- | --- | --- | --- |
| Title | Multi-Label Bengali article classification using ML-KNN algorithm and Neural Network | | |
| Author(s) Name | Wahiduzzaman Akanda; Ashraf Uddin | | |
| Contact Email(s) | dr.ashraf@aiub.edu | | |
| Published Journal Name |  | | |
| Type of Publication | Article | | |
| Volume |  | Issue |  |
| Publisher | IEEE | | |
| Publication Date | 27-2-21 | | |
| ISSN |  | | |
| DOI | https://doi.org/10.1109/ICICT4SD50815.2021.9396882 | | |
| URL | https://ieeexplore.ieee.org/document/9396882 | | |
| Other Related Info. |  | | |
|  | | | |

|  |  |
| --- | --- |
| Abstract |  |
| Multi-label classification is a very complex and critical task to solve in Natural Language Processing and Text Mining domain. Moreover, Bengali has limited resources to work with. The goal of this research is to overcome these constraints and provide a sophisticated and standard solution that will solve this problem for Bengali text. This research output can be utilized by any Bengali newspaper portals to improve their recommendation system as well as reduce manual labor of document tagging. In this work, we have utilized a large dataset that contains 4,16,289 news articles and 4,302 unique labels. These news articles are collected from one of the most popular Bengali newspapers of Bangladesh named Prothom Alo. The news articles span over seven years (2013 to 2019). These news articles are categorized into six categories named Sports, Technology, Economy, Entertainment, International, and State. This huge dataset helps us to build a supervised model using the ML-KNN algorithm and Neural Network. Furthermore, for the word embedding feature, we have utilized Count Vectorizer. We will also briefly discuss how different parameters like words per document, labels per category impact the result. | |