Maturity Level Rating System for Ebooks Using Longformer and Text-RCNN

Erik Miguel Go Celdran

University of San Carlos, Cebu, Philippines

Jastine Ouano Guzman

University of San Carlos, Cebu, Philippines

Abstract:

The growing accessibility of eBooks in the digital era calls for mechanisms that protect young readers from inappropriate content. This study fills this gap by proposing an automated maturity level rating system using natural language processing and machine learning models for eBooks. eBooks are classified into four maturity levels by the system: All Ages, Mild, Moderate, and Adult. A large dataset of full-text eBooks in English was obtained by scraping Book Cave. We split the texts into chunks to meet Longformer input limitations. A stratified sampling approach was applied to produce a balanced dataset by maturity level. The Longformer-TextRCNN model showed strong results with 76.2% accuracy, along with high precision, recall, and F1 scores. This research intends to offer a streamlined, standardized structure for assigning eBook ratings, it helps protect young readers from inappropriate content and increases the overall availability of digital literature. It also provides publishers and digital platforms with a standardized solution for ensuring consistency of content ratings across the industry.

Keywords:

eBooks, Maturity Level Rating, Longformer, TextRCNN, Natural Language Processing, Machine Learning