

ABSTRACT

In Indonesia, the Merdeka Curriculum's learning process mandates using books as essential teaching tools. The Ministry of Education, Culture, Research and Technology of the Republic of Indonesia regulates book quality standards to ensure book quality. To meet the book quality standards outlined in this regulation, it is necessary to have a tool that can analyze the text readability in the book. Assessing text readability can be accomplished by employing a readability formula and constructing a machine-learning model. In this study, we focus on analyzing the readability of Indonesian texts. We build readability models by applying readability formula and machine learning approach as a two-step process. The first step is classifying the text into two categories: complex and simple, and the next step is identifying the text difficulty level according to readability formula. The readability formula measures the grade level and reader age. The experimental results show that the machine learning model successfully classified the text with 83.1% accuracy. On the other hand, the result of the readability formula is that the readability age is higher than the reader's age. The result of the readability formula is the predicted readability age is higher than the intended reader age. We analyzed the readability result and found that Indonesian textbook has more complex linguistic feature than English textbook, therefore the predicted reader age is higher.

Index Terms: Text Readability Analysis, Readability Formula, Bi-LSTM, Indonesian