

ABSTRACT

COMPARISON OF NAIVE BAYES CLASSIFIER IN APPLICATION OF FLESCH KINCAID ON BSE-KEMENDIKBUD AS THE GOLD-STANDARD DATA SET

Sarah Fadhila Putri
17/411470/PA/17864

Books are very important as a learning medium between students and teachers. The ministry of education and culture (Kemendikbud) has provided books in electronic form that can be downloaded by every student and is well-known as Buku Sekolah Elektronik Kemendikbud (BSE-kemendikbud). The books that have been provided are classified and designated according to the respective classes from the grade 1-12 from Elementary to High School.

This research purpose is to measure the readability levels of the books that are provided by The Ministry of Education and Culture (Kemendikbud) using the Flesch Kincaid formula and with the help of a machine learning method. In this research, a book difficulty classification is carried out and classifies which books are Easy or Hard by using the Flesch Kincaid formula to find a score of the readability and this will do a comparison with the two models.

The first model uses the Flesch Kincaid formula to find the score and to label the data using a threshold that follows the provisions of the Flesch Kincaid Reading Ease Table formula. The second model will be modified to find a new threshold for the data labeling process. Both models will be classified using Gaussian Naive Bayes and Multinomial Naive Bayes which will be compared. The result of classification on the first model using Gaussian Naive Bayes showed the accuracy of 100% and using the Multinomial Naive Bayes is at 97.5%. The second model using Gaussian Naive Bayes is at the accuracy of 74.8% and using Multinomial Naive Bayes showed the accuracy at 75.6%.

Hyperparameter tuning is used in both models and both classifications in an attempt to optimize the accuracy obtained. With accuracy results on the first model using Gaussian Naive Bayes got the accuracy at 100% and first model using Multinomial Naive Bayes got the accuracy of 100%. The second model using Gaussian Naive Bayes got an accuracy of 98.3% and the second model using Multinomial Naive Bayes got an accuracy of 79.8%.

Keywords: BSE-Kemendikbud, Flesch Kincaid, Gaussian Naive Bayes, Multinomial Naive Bayes, Text Classification.