Sentiment Analysis terhadap Rencana Penghapusan Ujian Nasional menggunakan Algoritma *Support Vector Machine* dengan Pembobotan *TF-IDF*

Muhammad Lutfi Utama Putra1[1], Yuliant Sibaroni[2]

[1,2]Fakultas Informatika, Universitas Telkom, Bandung
[1]lutfiutama@students.telkomuniversity.ac.id, [2]yuliant@telkomuniversity.ac.id,

Abstract

During the second term of President Joko Widodo's administration through the Ministry of Education and Culture of the Republic of Indonesia, there were plans to remove the national exam and replace it with a minimum competency assessment and character survey. The new policy directive is planned to be realized in 2021. The national exam will be replaced with 3 assessment parameters, namely literacy, numeracy and character. In research [7], shows that the support vector machine algorithm has a higher accuracy value of 82.40% compared to Naive Bayes 78.90% and 72.90% decision tree. This study aims to find the best algorithm from twitter sentiment analysis in categorizing positive and negative sentiments that will be used as variables to calculate net brand reputation and can be used to make it easier for companies to systematically analyze and evaluate trends of a brand. Currently, it is not known how the public opinion towards the UN removal plan. The public can respond to this plan with positive, negative and even neutral opinions. This study uses the youtube platform as a source of data to explore public opinion. The TF-IDF weighted support vector machine algorithm was chosen because of its level of accuracy in various studies. Therefore, this study uses a support vector machine algorithm with TF-IDF weighting to determine public opinion towards the plan to remove the national exam. The results of sentiment analysis from 2422 comments with 10% testing data, namely 243 labels consisting of 190 positive, 28 neutral and 25 negative, can measure the accuracy of the support vector machine algorithm with TF-IDF weighting of 81.84% using a linear kernel with C = 1.0 and test size 0.1, 63.50% using the RBF kernel with a test size of 0.5 and 64.12% using a polynomial kernel with a test size of 0.2. Based on 13221 unlabeled data, the SVM model with the best accuracy, namely 81.84% succeeded in predicting 8346 positive labels or 62.7% of the public agreed that the National Exam was removed, 2382 negative labels or 17.9% of the public rejected the National Exam being removed, 2593 neutral labels or 19.5% of the community neither agreed nor did not reject UN to removed.

Keywords: sentiment analysis, national exam, support vector machine, youtube, TF-IDF