

## ABSTRAK

Pandemi COVID-19 menjadi permasalahan di Indonesia dan menyebabkan dampak, khususnya bidang pendidikan. Bidang pendidikan menjadi salah satu bidang yang sangat terdampak akibat pandemi tersebut. Oleh karena itu pemerintah memiliki sebuah solusi yaitu pembelajaran secara daring atau pembelajaran jarak jauh. Pembelajaran daring diharapkan dapat menggantikan pembelajaran luring demi mengurangi aktivitas tatap muka. Pembelajaran daring tersebut ternyata juga memiliki dampak positif ataupun negatif. Hal tersebut dapat dilihat dari media sosial Twitter.

Media sosial Twitter memiliki data *tweet* yang berasal dari masyarakat, khususnya Indonesia. Berdasarkan *tweet* tersebut, dapat dilakukan penelitian untuk mengetahui sentimen terhadap pembelajaran daring. Data *tweet* didapatkan melalui *scrapping* dengan aplikasi *Spyder Python*. Data yang diambil adalah data *tweet* bulan Maret 2020 sampai dengan Juni 2020 dengan total data kurang lebih 5236 *tweet*. Data akan melalui tahap *preprocessing*, kemudian terjemahkan dokumen, pemberian label menggunakan *Vader Lexicon*, pembagian data menggunakan *k-fold cross validation*, penghitungan bobot TF-IDF, klasifikasi dengan *Support Vector Machine*, dan terakhir mengevaluasi klasifikasi SVM dengan menggunakan *confusion matrix*.

Berdasarkan hasil pengujian yang telah dilakukan, *Vader Lexicon* dapat memberikan label sentimen dari setiap *tweet*. Klasifikasi dilakukan dengan nilai  $k$  (3,5,10) dan kernel (*linear*, *rbf*, *polynomial*). Berdasarkan model evaluasi *confusion matrix* tersebut dapat dihasilkan bahwa hasil klasifikasi dengan menggunakan klasifikasi *Support Vector Machine* (SVM) didapatkan hasil akurasi tertinggi sebesar 74.51% dengan nilai  $k = 10$ , kernel = *rbf* dengan nilai recall = 80.94 % dan nilai *precision* = 77.17%.

Kata Kunci : *Tweet*, Klasifikasi, *Vader Lexicon*, *Support Vector Machine*.

## ABSTRACT

The COVID-19 pandemic has become a problem in Indonesia and has had an impact, especially in the education sector. The education sector is one of the areas that has been most affected by the pandemic. Therefore, the government has a solution, namely online learning or distance learning. Online learning is expected to replace offline learning in order to reduce face-to-face activities. Online learning also has a positive or negative impact. This can be seen from the social media Twitter.

Twitter social media has tweet data from the public, especially Indonesia. Based on these tweets, research can be carried out to determine sentiments towards online learning. Tweet data is obtained through the scraping with the Spyder Python application. The data taken is tweet data from March 2020 to June 2020 with a total data of approximately 5236 tweets. The data will go through the preprocessing stage, then translate the document, label it using Vader Lexicon, share data using k-fold cross validation, calculate TF-IDF weights, classify with Support Vector Machine, and finally evaluate SVM classification using confusion matrix. Based on the results of the tests that have been carried out, Vader Lexicon can provide a sentiment label for each tweet. Classification is done with the value of k (3,5,10) and kernel (linear, rbf, polynomial). Based on the confusion matrix evaluation model, it can be concluded that the classification results using the Support Vector Machine (SVM) classification obtained the highest accuracy of 74.51% with a value of k = 10, kernel = rbf with a recall value = 80.94% and a precision value = 77.17%.

Keywords: Tweet, Classification, Vader Lexicon, Support Vector Machine.