ABSTRACT

In this analysis, the method used is the K-Nearest Neighbor classification method and the Logistic Regression classification method with data taken on the twitter application. This study examines the level of accuracy in public sentiment regarding the Covid-19 vaccination with positive and negative labels. The AUC value in the KNN algorithm with TextBlob labeling is 0.765 and 0.768 for VaderSentiment labeling, both of which are included in the fair classification criteria. Meanwhile, the Logistic Regression algorithm produces an accuracy of 84.97% with a ratio of 90:10 for TextBlob labeling, while VaderSentiment labeling with a ratio of 90:10 produces an accuracy of 85.22%. Both algorithms are validated using K-Fold Cross Validation with a number of folds of 10. The comparison results obtained when evaluating the confusion matrix show that the Logistic Regression algorithm with VaderSentiment labeling has the highest accuracy value compared to the K-Nearest Neighbor algorithm with TextBlob and VaderSentiment labeling .

Keywords: Covid-19 Vaccination, K-Nearest Neighbor, Logistic Regression, Sentiment Analysis, TextBlob, VaderSentiment