

**KLASIFIKASI SENTIMEN DATA TIDAK SEIMBANG MENGGUNAKAN
ALGORITMA SMOTE DAN *K-NEAREST NEIGHBOR* PADA ULASAN
PENGGUNA APLIKASI PEDULILINDUNGI**

SHEILA GABRIELA BARUS

ABSTRAK

Salah satu penanganan pemerintah dalam mengatasi penyebaran Covid-19 yang terjadi di Indonesia yaitu dengan membuat sebuah aplikasi yaitu aplikasi PeduliLindungi. Aplikasi ini berfungsi dalam melacak dan memantau penyebaran Covid-19, oleh karena itu banyak masyarakat Indonesia yang harus mempunyai aplikasi ini. Banyak juga ulasan yang diberikan pada aplikasi ini, dari komentar yang positif hingga komentar negatif. Ulasan tersebut yang menjadi data dalam penelitian ini untuk mengetahui hasil sentimen masyarakat dan menguji klasifikasi algoritma *K-Nearest Neighbor*. Pengumpulan data dilakukan dengan *scraping* di google play menggunakan bahasa pemrograman *Python*, dimana data yang diperoleh mendapatkan 750 label negatif dan 250 label positif. Sehingga data yang tidak seimbang ini harus diseimbangkan dengan teknik *undersampling* dan *oversampling* SMOTE. Oleh karena itu, penelitian ini dilakukan tiga data yang berbeda jumlah yaitu dari data yang tidak seimbang, data yang sudah di *undersampling* dan data yang sudah di *oversampling* dengan SMOTE. Hasil dari ketiga percobaan tersebut diperoleh nilai terbaik menggunakan teknik SMOTE pada $K = 1$ dengan nilai akurasi sebesar 0.9766, nilai presisi sebesar 0.9691, nilai *F1 score* 0.9781, nilai spesifisitas sebesar 0.9645, dan nilai sensitivitas sebesar 0.9874.

Kata Kunci : Sentimen, *K-Nearest Neighbor*, SMOTE, PeduliLindungi

**CLASSIFICATION OF UNBALANCED DATA SENTIMENTS USING
SMOTE ALGORITHM AND K-NEAREST NEIGHBOR ON USER REVIEWS
OF CARE APPLICATIONS**

SHEILA GABRIELA BARUS

ABSTRACT

One of the government's methods in dealing with the spread of Covid-19 that occurred in Indonesia is to create an application, namely the PeduliLindungi application. This application functions in tracking and monitoring the spread of Covid-19, therefore many Indonesian people must have this application. Many reviews are also given on this application, from positive comments to negative comments. These reviews are used as data in this study to determine the results of community sentiment and to test the classification of the K-Nearest Neighbor algorithm. Data collection was done by scraping on google play using the Python programming language, where the data obtained got 750 negative labels and 250 positive labels. So this unbalanced data must be balanced with SMOTE undersampling and oversampling techniques. Therefore, this study carried out three experiments, namely from unbalanced data, data that had been undersampled and data that had been oversampled with SMOTE. The results of the three experiments obtained the best value using the SMOTE technique at $K = 1$ with an accuracy value of 0.9766, a precision value of 0.9691, an F1 score of 0.9781, a specificity value of 0.9645, and a sensitivity value of 0.9874.

Keywords : *Sentiment, K-Nearest Neighbor, SMOTE, PeduliLindungi*