

ABSTRACTS

The research aims to evaluate the performance of two classification algorithms, namely Support Vector Machine (SVM) and Naive Bayes Classifier (NBC), in two different scenarios: balanced and unbalanced data. The research findings reveal that improved preprocessing significantly enhances sentiment classification performance, regardless of whether the data is balanced or unbalanced. In the sentiment analysis implementation using balanced data as training data, consisting of 2685 training data and 672 test data, the SVM algorithm achieves an accuracy rate of 69%, precision of 69%, recall of 69%, and F-measure of 69%. The NBC algorithm also yields a similar accuracy rate of 70%, with precision, recall, and F-measure all at 70%. When sentiment analysis is applied to unbalanced data using 5038 training data and 1260 test data, the SVM algorithm achieves an accuracy rate of 76%, with precision of 70%, recall of 69%, and F-measure of 69%. The NBC algorithm demonstrates improved performance with an accuracy rate of 78%, precision of 72%, recall of 72%, and F-measure of 72%. The research results highlight differences in algorithm performance when confronted with balanced and unbalanced data, with a significant improvement in accuracy and other sentiment metrics when using unbalanced data.

Keywords: Classification Algorithms; Sentiment Analysis; Naive Bayes Classifier; Preprocessing; and Support Vector Machine.

ABSTRAKS

Penelitian bertujuan untuk mengevaluasi kinerja dua algoritma klasifikasi, yaitu Support Vector Machine (SVM) dan Naive Bayes Classifier (NBC), dalam dua skenario berbeda: data seimbang dan data tidak seimbang. Hasil penelitian mengungkapkan bahwa proses preprocessing yang lebih baik secara signifikan meningkatkan kinerja klasifikasi sentimen, terlepas dari apakah data seimbang atau tidak. Pada implementasi analisis sentimen dengan menggunakan data yang seimbang sebagai data latih, terdiri dari 2685 data latih dan 672 data uji, algoritma SVM mencapai tingkat akurasi sebesar 69%, presisi 69%, recall 69%, dan F-measure 69%. Algoritma NBC juga menghasilkan tingkat akurasi yang serupa, yaitu 70%, dengan presisi 70%, recall 70%, dan F-measure 70%. Ketika analisis sentimen diterapkan pada data yang tidak seimbang dengan menggunakan data latih sebanyak 5038 data dan data uji sebanyak 1260 data, algoritma SVM mencapai tingkat akurasi sebesar 76%, dengan presisi 70%, recall 69%, dan F-measure 69%. Algoritma NBC menunjukkan peningkatan kinerja dengan tingkat akurasi sebesar 78%, dengan presisi 72%, recall 72%, dan F-measure 72%. Hasil penelitian ini menunjukkan perbedaan dalam kinerja algoritma ketika dihadapkan pada data yang seimbang dan tidak seimbang, dengan peningkatan yang signifikan pada akurasi dan metrik sentimen lainnya ketika menggunakan data yang tidak seimbang.

Kata Kunci: Algoritma Klasifikasi; Analisis Sentimen; *Naive Bayes Classifier*; *Preprocessing*; dan *Support Vector Machine*.