

ABSTRAK

PERBANDINGAN PELABELAN INDONESIA ROBERTA BASE SENTIMENT CLASSIFIER DAN PELABELAN FUZZY C-MEANS UNTUK KLASIFIKASI SENTIMEN DENGAN SVM PADA LABEL APLIKASI AKULAKU

Fransiska Febyanti

Universitas Sanata Dharma

2024

Penelitian ini membandingkan dua metode pelabelan data untuk klasifikasi sentimen pada aplikasi Akulaku: *Indonesian Roberta Base Sentiment Classifier* dan *Fuzzy C-Means*. Tujuan dari penelitian ini adalah untuk mengevaluasi performa kedua metode dalam mengklasifikasikan sentimen komentar pengguna aplikasi. Proses penelitian melibatkan pengumpulan data komentar pengguna, *pre-processing* teks, pelabelan data menggunakan kedua metode, dan analisis hasil menggunakan *Support Vector Machine* (SVM). Hasil penelitian menunjukkan bahwa metode pelabelan dengan *Fuzzy C-Means* memberikan akurasi yang lebih tinggi dibandingkan dengan metode *Indonesian Roberta Base*. Pada pengujian dengan nilai $K = 7$, metode *Fuzzy C-Means* dengan kernel RBF, parameter $C = 100$ dan $\gamma = 0.1$ menghasilkan *accuracy* sebesar 96.89%, *precision* 96.54%, *recall* 96.81%, dan *F1-score* 96.67%. Sementara itu, metode *Roberta* dengan kernel RBF, parameter $C = 1$ dan $\gamma = 1$ menghasilkan *accuracy* sebesar 91.71%, *precision* 91.94%, *recall* 91.11%, dan *F1-score* 91.45%.

Kata kunci : Analisis Sentimen, *Indonesian Roberta Base Sentiment Classifier*, *Fuzzy C-Means*, *Support Vector Machine* (SVM).

ABSTRACT

COMPARISON OF INDONESIAN ROBERTA BASE SENTIMENT CLASSIFIER LABELING AND FUZZY C-MEANS LABELING FOR SENTIMENT CLASSIFICATION WITH SVM ON AKULAKU APP LABELS

Fransiska Febyanti

Universitas Sanata Dharma

2024

This research compares two data labeling methods for sentiment classification on the Akulaku application: Indonesian Roberta Base Sentiment Classifier and Fuzzy C-Means. The purpose of this research is to evaluate the performance of both methods in classifying the sentiment of app users' comments. The research process involves user comment data collection, text pre-processing, data labeling using both methods, and result analysis using Support Vector Machine (SVM). The results show that the labeling method with Fuzzy C-Means provides higher accuracy compared to the Indonesian Roberta Base method. In testing with a value of $K = 7$, the Fuzzy C-Means method with RBF kernel, parameters $C = 100$ and $\gamma = 0.1$ resulted in an accuracy of 96.89%, precision 96.54%, recall 96.81%, and F1-score 96.67%. Meanwhile, the IRbSC method with RBF kernel, parameter $C = 1$ and $\gamma = 1$ resulted in an accuracy of 91.71%, precision 91.94%, recall 91.11%, and F1-score 91.45%.

Keyword : *Sentiment Analysis, Indonesian Roberta Base Sentiment Classifier, Fuzzy C-Means, Support Vector Machine (SVM).*