

ABSTRACT

Sentiment analysis is a computational study that examines opinions, feelings, and emotions in text, using techniques such as natural language processing, computational linguistics, and text mining. The aim is to classify the polarity of the text with algorithms such as Naive Bayes and Logistic Regression, which have shown significant accuracy in various studies. This study specifically reviewed comments on YouTube about the debate between presidential and vice-presidential candidates in the 2024 general election in the Republic of Indonesia. Public opinion has an important influence on the electability of prospective leader pairs. This final project aims to develop applications that can compare the effectiveness of the two algorithms in classifying sentiments, using metrics such as accuracy, precision, and recall. The selection of these two algorithms is based on the balanced accuracy of related research and the use of the principles of probability and statistics. This final project uses 108,867 comments taken from YouTube, then processed through the Text Preprocessing stage involving case folding, tokenizing, cleaning, stopword removal, and stemming to produce clean and ready-to-use data. The test results show that the two algorithms are able to classify sentiments into positive, negative, and neutral classes, although there are differences in the performance of each algorithm. In 10 iterations, the Naive Bayes algorithm obtained an accuracy of 85.17% with an error rate of 0.039, while Logistic Regression obtained an accuracy of 90.67% and an error rate of 0.053. The application developed is not only able to conduct sentiment analysis on mass data but also provides features for testing new comments individually, displaying the results of the analysis in the form of line diagrams and processed data collections.

Keywords: *Sentiment Analysis, Naive Bayes, Logistics Regression, YouTube comments.*

ABSTRAK

Analisis sentimen adalah studi komputasional yang meneliti opini, perasaan, dan emosi dalam teks, menggunakan teknik seperti pengolahan bahasa alami, komputasi linguistik, dan *text mining*. Tujuannya adalah mengklasifikasikan polaritas teks dengan algoritma seperti *Naive Bayes* dan *Logistic Regression*, yang telah menunjukkan akurasi yang signifikan dalam berbagai penelitian. Studi ini khususnya mengkaji komentar pada *YouTube* tentang debat Calon Presiden dan Calon Wakil Presiden pada pemilihan umum 2024 di Republik Indonesia. Opini publik memiliki pengaruh penting terhadap elektabilitas pasangan calon pemimpin, pada Tugas Akhir ini bertujuan mengembangkan aplikasi yang dapat membandingkan efektivitas kedua algoritma dalam mengklasifikasikan sentimen, dengan menggunakan metrik seperti akurasi, presisi, dan *recall*. Pemilihan kedua algoritma ini didasarkan pada akurasi penelitian terkait yang seimbang dan penggunaan prinsip probabilitas, serta statistika. Tugas Akhir ini menggunakan data sebanyak 108.867 komentar yang diambil dari *YouTube*, yang kemudian diproses melalui tahap *text preprocessing* yang melibatkan *case folding*, *tokenizing*, *cleaning*, *stopword removal*, dan *stemming* untuk menghasilkan data yang bersih dan siap digunakan. Hasil pengujian menunjukkan bahwa kedua algoritma mampu mengklasifikasikan sentimen ke dalam kelas positif, negatif, dan netral, meskipun terdapat perbedaan dalam performa masing-masing algoritma. Dalam 10 kali iterasi diperoleh akurasi sebesar 85,17% untuk algoritma *Naive Bayes* dengan 0,039 *error rate*, sedangkan untuk *Logistic Regression* diperoleh akurasi sebesar 90,67% dan *error rate* sebesar 0,053. Aplikasi yang dikembangkan tidak hanya mampu melakukan analisis sentimen pada data massal, tetapi juga menyediakan fitur untuk pengujian komentar baru secara individu, menampilkan hasil analisis dalam bentuk diagram garis dan himpunan data yang telah diproses.

Kata Kunci: Analisis Sentimen, *Naive Bayes*, *Logistic Regression*, Komentar *YouTube*.