ABSTRAK

Tidak sedikit masyarakat indonesia yang mengutarakan opininya terkait penilaian dan kebijakan tokoh-tokoh politik. maka dari itu, penelitian ini akan mengklasifikasikan sentimen positif dan negatif menggunakan model algoritma Naïve Bayes dan aplikasi rapidminer. Dalam penelitian ini menggunakan metode pengembangan SEMMA. Pada pengumpulan data menggunakan teknik *scraping* menghasilkan data sebanyak 25.000 dengan rentang waktu dari awal tahun 2018 sampai akhir tahun 2023 lalu dilanjutkan dengan tahap proses pengolahan data sampai data siap untuk diolah oleh model Klasifikasi Naïve Bayes. Akan tetapi dalam penelitian ini model algoritma *machine learning* Naïve Bayes dibagi menjadi tiga model Klasifikasi yaitu model Klasifikasi yang melewati proses *Stemming* menggunakan algoritma Sastrawi, model Klasifikasi yang melewati proses *Stemming* menggunakan algoritma Nazief & Adriani, dan model Klasifikasi yang tidak melewati proses *Stemming*. Hasil yang didapat oleh analisis sentimen twitter menggunakan model Klasifikasi dengan proses *Stemming* menggunakan algoritma Sastrawi menghasilkan akurasi 82%, presisi 94%, *recall* 83%, dan *F1-score* 88%. Hasil yang didapat oleh analisis sentimen twitter dengan proses *Stemming* menggunakan algoritma Nazief & Adriani menghasilkan akurasi 83%, presisi 94%, *recall* 84%, dan *F1-score* 89%. Hasil yang didapat oleh analisis sentimen twitter tanpa proses *Stemming* menghasilkan akurasi 88%, presisi 95%, *recall* 90%, dan *F1-score* 93%. Hasil yang didapat oleh analisis sentimen media online dengan proses *Stemming* menggunakan algoritma Sastrawi menghasilkan akurasi 86%, presisi 88%, *recall* 95%, dan *F1-score* 91%. Hasil yang didapat oleh analisis sentimen media online dengan proses *Stemming* menggunakan algoritma Nazief & Adrian menghasilkan akurasi 87%, presisi 89%, *recall* 95%, dan *F1-score* 92%. Hasil yang didapat oleh analisis sentimen media online tanpa proses *Stemming* menghasilkan akurasi 86%, presisi 87%, *recall* 95%, dan *F1-score* 91%. Berdasarkan hasil tersebut model Klasifikasi Naïve Bayes yang tidak menggunakan proses *Stemming* memiliki kinerja paling baik.

Kata kunci : Analisis Sentimen dan *Preprocessing text.*

ABSTRACT

Many Indonesians express their opinions on the assessment and policies of political figures. This research classifies positive and negative sentiments using the Naïve Bayes algorithm model and the RapidMiner application. The study utilizes the SEMMA development method. Data collection through scraping techniques yielded 25,000 pieces of data spanning from early 2018 to the end of 2023. This was followed by the data processing stage and then processed using the Naïve Bayes Classification model. In this research, the Naïve Bayes machine learning algorithm model is divided into three classification models: a classification model that undergoes Stemming using the Sastrawi algorithm, a classification model that undergoes Stemming using the Nazief & Adriani algorithm, and a classification model that does not undergo Stemming. The sentiment analysis of Twitter using the classification model with Stemming using the Sastrawi algorithm resulted in an accuracy of 82%, precision of 94%, recall of 83%, and an F1-score of 88%. The sentiment analysis of Twitter with Stemming using the Nazief & Adriani algorithm produced an accuracy of 83%, precision of 94%, recall of 84%, and an F1-score of 89%. The sentiment analysis of Twitter without Stemming achieved an accuracy of 88%, precision of 95%, recall of 90%, and an F1-score of 93%. The sentiment analysis of online media with Stemming using the Sastrawi algorithm resulted in an accuracy of 86%, precision of 88%, recall of 95%, and an F1-score of 91%. The sentiment analysis of online media with Stemming using the Nazief & Adriani algorithm produced an accuracy of 87%, precision of 89%, recall of 95%, and an F1-score of 92%. The sentiment analysis of online media without Stemming achieved an accuracy of 86%, precision of 87%, recall of 95%, and an F1-score of 91%. In conclusion, the results indicate that the Naïve Bayes Classification model without Stemming performs the best.

Keywords: *Sentiment Analysis, Text Preprocessing.*