ABSTRAK

Opini publik terhadap tokoh politik dan kebijakan memiliki dampak besar pada pemilihan umum. Analisis sentimen, sebagai metode untuk memahami opini dan emosi dalam teks, memerlukan langkah *Text Preprocessing* untuk meningkatkan kualitas data. Namun, data teks seringkali memiliki masalah seperti kata-kata tidak relevan, tanda baca tidak konsisten, dan bahasa ambigu. Kondisi ini dapat mempengaruhi akurasi analisis sentimen, terutama dalam konteks analisis terhadap tokoh politik. Penelitian ini fokus pada pengaruh teknik *preprocessing* khusus, yaitu *Phrase Detection* dengan *N-Gram*, terhadap analisis sentimen terhadap tokoh politik. Hasil penelitian menunjukkan bahwa penggunaan *Bigram* dalam *Phrase Detection* memberikan hasil yang lebih signifikan dibandingkan dengan *Trigram* dan *Unigram* pada sebagian besar tokoh politik di *Twitter* dengan nilai akurasi tertinggi sebesar 88,23%. Analisis sentimen artikel media *online* juga menunjukkan variasi hasil tergantung pada jenis *N-Gram*. Penelitian ini memberikan kontribusi baru dalam pemahaman terhadap pengaruh teknik *preprocessing* khusus terhadap analisis sentimen politik, dengan fokus pada media *online* dan *Twitter*.

Kata Kunci: Analisis sentimen, *Text Preprocessing*, *N-Gram*, politik, opini, media *online*, *Twitter*

ABSTRACT

Public opinion of political figures and policy has a big impact on general elections. Sentiment analysis, as a method to comprehend opinion and emotion in texts, requires the step of text preprocessing to improve data quality. However, textual data often contain problems such as irrelevant words, inconsistent punctuations, and ambiguous expressions. Such conditions can influence the accuracy of sentiment analysis, particularly in the context of analysis of political figures. This research focuses on the influences of specific preprocessing techniques, namely Phrase Detection with N-Gram, on the sentiment analysis for political figures. The results of this research indicate that the use of Bigram in Phrase Detection provides more significant results compared with Trigram and Unigram in most political figures at Twitter with the highest accuracy score of 88,23%. Sentiment analysis of articles in online media also indicates various results depending on the type of N-Gram. This research provides a new contribution to understanding the influence of specific preprocessing techniques on political sentiment analysis, focusing on online media and Twitter.

Keywords: Sentiment analysis, Text Preprocessing, N-Gram, politics, opinion, online media, Twitter