

INTISARI

Terdapat beberapa tantangan dalam memproses data teks dari *microblog* atau suatu *website*. Tantangan terbesar yang seringkali ditemukan adalah struktur kata dalam teks yang tidak formal, seperti penggunaan singkatan, mengganti huruf dengan angka, dan penggunaan tanda baca yang tidak tepat. Kesalahan pada suatu tulisan atau dokumen teks dapat menyebabkan sulitnya proses *data preprocessing* dan berkurangnya akurasi dari proses lanjutan, salah satunya klasifikasi sentimen.

Library Maleo adalah salah satu modul yang dapat membantu proses *text preprocessing* menjadi lebih mudah dengan beragam fitur yang dimiliki, seperti *scanner*; menghapus *hyperlink*, tanda baca, *stop word*, *emoticon*, dan lainnya; mengekstrak *hashtags* dan harga dari teks; mengubah *email*, nomor telepon dan tanggal menjadi <TAG>; mengubah *slang* atau OOV dalam Bahasa Indonesia menjadi baku; dan mengubah kata menjadi angka.

Penelitian ini bertujuan untuk mengetahui apakah Maleo dapat berpengaruh atau bahkan meningkatkan kinerja algoritma *text preprocessing* dan klasifikasi sentimen saat menangani teks Bahasa Indonesia dengan studi kasus ulasan daring produk otomotif Toyota Avanza pada situs Seraya Motor. Penulis membandingkan parameter akurasi, *recall*, presisi dan F1-Score dari kinerja algoritma *text preprocessing* dan klasifikasi saat menggunakan *library Maleo* dan tanpa menggunakan *library Maleo*.

Kata kunci: *text preprocessing*, klasifikasi sentimen, *confusion matrix*, *NLP Library*, *Library Maleo*

ABSTRACT

There are several challenges in processing text data from a microblog or a website. The biggest challenge that is often encountered is the informal word structure of the text, such as using abbreviations, replacing letters with numbers, and using inappropriate punctuation. Errors in writing or text documents can cause difficulty in data preprocessing and reduce the accuracy of further text processing, one of which is sentiment classification or sentiment analysis.

Maleo library is a module that can make the text preprocessing process easier with its various features, such as text scanner; remove hyperlinks, punctuation marks, stop words, emoticons, and others; extract hashtags and prices from text; change email, phone number and date to <TAG>; changing slang or OOV in Indonesian to standard words; and converts words to numbers

This study aims to determine whether Maleo can affect or even improve the performance of the text preprocessing algorithm and sentiment classification when handling Indonesian text with a case study of online review of Toyota Avanza products on the Seraya Motor website. The author compares the parameters of accuracy, recall, precision and F1-Score of the performance of the text preprocessing algorithm and classification when using the Maleo library and without using the Maleo library.

Keywords: text preprocessing, sentiment analysis, confusion matrix, NLP Library, Library Maleo.