

## SARI

Dalam komunikasi sehari-hari, ada dua macam model bahasa yang banyak digunakan masyarakat di Indonesia, yaitu bahasa formal dan bahasa tidak formal. Bahasa formal merupakan bahasa yang mengikuti kaidah bahasa, sedangkan bahasa yang tidak formal adalah bahasa yang keluar dari kaidah baku bahasa. Namun demikian, masyarakat lebih sering menggunakan bahasa non-formal untuk komunikasi sehari-hari. Penggunaan bahasa tersebut digunakan di antaranya dengan alasan untuk memperlancar komunikasi dan menyederhanakan tulisan pesan. Pengolahan bahasa alami adalah salah satu bidang ilmu yang menangani bentuk bahasa tersebut untuk berinteraksi dengan komputer. Dalam pengolahan bahasa alami, kemunculan kata tidak standar seperti di atas sangat mempengaruhi dalam proses pengolahan kata. Oleh karena itu, ada sebuah tahapan yang disebut dengan *preprocessing*.

Berdasarkan permasalahan ini, dibuatlah proses secara otomatis untuk menangani perubahan kata yang tidak standar menjadi standar. Maka, diusulkan sebuah solusi untuk menangani permasalahan tersebut dengan algoritma *Levenshtein Distance*. Perbaikan bahasa meliputi *slangword*, singkatan dan salah ejaan, *slangword* dan singkatan menggunakan kamus untuk memperbaiki kata sedangkan algoritma *Levenshtein Distance* diterapkan pada perbaikan salah ejaan. Algoritma *Levenshtein Distance* mencari kata terdekat pada kamus kata standar, kamus kata standar diambil dari kata-kata yang sering digunakan dalam percakapan sehari-hari. Akurasi yang diperoleh dalam penelitian ini dengan menggunakan algoritma *Levenshtein Distance* adalah sebesar 66,89% dengan jumlah data uji sebanyak 106 kata.

Kata kunci: *slangword*, singkatan, salah ejaan, *Levenshtein Distance*, bahasa non-formal