

## ABSTRAK

### KOREKSI KESALAHAN PENGETIKAN KATA DALAM BAHASA INDONESIA MENGGUNAKAN *MACHINE LEARNING*

Oleh  
Alma Alfiatul Inayah  
17102050

Kesalahan pemakaian ejaan seringkali ditemukan dalam sebuah tulisan, yang diakibatkan oleh kesalahan pengetikan. Kesalahan ketik menyebabkan suatu kata bermakna ambigu atau tidak sesuai KBBI. Kesalahan dalam penulisan sering kali berdampak pada karya tulis yang mengalami revisi dikarenakan penulisan yang tidak sesuai dengan EYD (Ejaan Yang Disempurnakan). Banyak penelitian menghasilkan sistem koreksi otomatis yang dapat mengoreksi kesalahan pengetikan. *Levenshtein Distance*, *Soundex*, *Recurrent Neural Network*, dan *FastText* adalah algoritma yang digunakan untuk mengoreksi kesalahan pengetikan. Perbedaan algoritmanya terletak pada cara kerjanya. Cara kerja *Levenshtein Distance* adalah dengan menghitung jarak terdekat antara dua buah string. *Soundex* mengoreksi kesalahan pengetikan berdasarkan bunyinya. *Recurrent Neural Network* akan mengoreksi kesalahan pengetikan dengan memprediksi kata yang salah ejaan/pengetikannya. Sedangkan *FastText* akan mengoreksi kata yang salah berdasarkan kemiripan vektornya. Dari hal tersebut, tujuan dari penelitian ini adalah untuk membandingkan akurasi dan kecepatan pemrosesan keempat algoritma tersebut dalam mengoreksi kesalahan pengetikan sesuai dengan konteks kalimat. Sebagai bahan pengujian, dataset yang digunakan untuk membangun model ada 2, yaitu kata dasar dalam KBBI sebanyak 36.517 kata dan kumpulan korpus novel. Berdasarkan pengujian menggunakan sampel 100 kalimat dan 1000 kalimat. Algoritma dengan tingkat akurasi terbaik dihasilkan pada model *FastText* sebesar 44% untuk 100 kalimat dan 35% untuk 1000 kalimat, dengan kecepatan pemrosesan 0,00771513 detik per kalimat.

**Kata kunci :** *Koreksi Pengetikan, Levenshtein Distance, Soundex, Recurrent Neural Network, Fasttext.*