

## Abstrak

Gaya penulisan informal dan penulisan kata yang tidak tepat sering kali menimbulkan dampak negatif pada beberapa penelitian *Natural Language Processing*(NLP), untuk itu diperlukan proses normalisasi yang berperan merubah bentuk data yang tidak terstruktur menjadi lebih terstruktur dengan standar yang telah ditentukan, normalisasi dinilai dapat mengurangi dampak dari gaya penulisan informal dan meningkatkan performa NLP.

Pada penelitian ini akan dilakukan proses normalisasi dengan pendekatan berbasis mesin translasi, menggunakan arsitektur *Sequence to Sequence* dengan jaringan lstm dan mekanisme *attention*, dataset yang digunakan dikumpulkan dari dataset penelitian mesin translasi dengan target bahasa indonesia, penggalan komentar dari media sosial Twitter, dan melalui kuisioner kalimat sehari-hari. Hasil keluaran sistem diharapkan dapat berupa kalimat yang baku dan dapat di- proses dengan baik oleh semua permodelan NLP.

Pada penelitian ini didapatkan nilai rata-rata akurasi dengan algoritma BLEU *score* sebesar 27.44% untuk data uji yang seluruh katanya dikenali oleh kamus tokenisasi, dan 18.49% untuk data uji yang mengandung kosa kata diluar dari kamus tokenisasi. Berdasarkan nilai akurasi tersebut, sistem yang dibangun dinilai belum cukup cakap untuk di implementasikan secara langsung dan masih membutuhkan banyak pengembangan dan pelatihan lebih lanjut.

**Kata Kunci:** Normalisasi, kalimat tidak baku, bahasa indonesia, *Sequence to Sequence*.