

ABSTRAK

Tingginya aktivitas masyarakat dalam membahas pemilihan Gubernur melalui media sosial menghasilkan data komentar dalam jumlah besar, namun komentar tersebut sering menggunakan bahasa informal, bahasa sehari-hari, singkatan, serta bercampur dengan bahasa daerah dan dialek lokal yang sulit dipahami. Hal ini menghambat pemrosesan data komentar untuk keperluan analisis atau tujuan lainnya.

Proses normalisasi manual membutuhkan waktu dan sumber daya yang sangat banyak, terutama jika data yang diolah berjumlah besar. Normalisasi secara manual juga rentan terhadap inkonsistensi dan kesalahan manusia. Jumlah data komentar di media sosial yang terus meningkat membuat normalisasi manual semakin tidak mungkin dan tidak efisien untuk dilakukan, sehingga diperlukan solusi otomatisasi.

Sistem normalisasi teks otomatis dikembangkan menggunakan pendekatan *Phrase-Based Statistical Machine Translation* dengan memanfaatkan Moses. Dataset korpus paralel dibangun dari 31.889 pasangan kalimat informal-formal, sedangkan korpus monolingual terdiri dari 1.613.381 kalimat yang diambil dari Wikipedia. Model dievaluasi menggunakan metrik *BLEU* untuk mengukur kualitas hasil normalisasi.

Model terbaik mencapai skor *BLEU* 82,16 pada data test dan 81,04 pada data validasi, berhasil mengenali berbagai pola bahasa informal seperti singkatan tidak baku, kata berulang dengan angka, dan bahasa gaul. Namun, sistem memiliki keterbatasan terhadap kemampuan penanganan *Out-Of-Vocabulary*.

Kata Kunci: normalisasi teks, *PBSMT*, Moses, media sosial, pemilihan gubernur