

---

## Abstrak

*Stemming* merupakan bagian penting dari proses *preprocessing* dalam pengerjaan analisis sentimen. Pada penelitian sebelumnya, permasalahan *stemming* untuk *overstemming* dan *understemming* belum dapat diselesaikan menggunakan algoritma *stemming* Bahasa Indonesia. Studi ini menganalisis dan mengukur akurasi dari pengaruh *stemming* Bahasa Indonesia melalui analisis sentimen dari data media *twitter*. Selanjutnya studi akan membandingkan pengaruh akurasi antara tiga kondisi, yaitu proses analisis sentimen tanpa menggunakan *stemming*, proses analisis sentimen dengan menggunakan proses *stemming* dengan algoritma Nazief Adriani, dan proses sentimen menggunakan *stemming* Sastrawi. Penelitian ini mengumpulkan 2500 tweets dari Juli 2018 hingga Desember 2018 dengan menggunakan kata kunci Gojek. Lalu, kami memproses *tweet* dan memberikan pelabelan secara manual sebelum dilakukan praproses dan membandingkan pengaruh tanpa *stemming*, dan algoritma yang telah disebutkan diatas, lalu ketiganya akan diekstraksi fitur menggunakan *unigram*, *bigram* dan *unigram bigram*, kemudian akan dikelompokkan menggunakan algoritma SVM. Lalu dilakukan pengujian melalui *confusion matrix* antara kelompok data tersebut untuk mengevaluasi akurasi. Hasil dalam penelitian ini menunjukkan *stemming* Sastrawi cenderung lebih meningkatkan akurasi sebanyak 0.604% dibanding tidak menggunakan *stemming* maupun menggunakan *stemming* Nazief Adriani. *Stemming* Sastrawi juga dapat mengatasi permasalahan *overstemming* dan *understemming*, namun belum bisa mengatasi permasalahan *word sense disambiguation*

**Kata kunci :** Analisis Sentimen, *Stemming*, Ekstraksi Fitur, SVM.