

## Abstrak

Media sosial adalah forum tempat pengguna dapat berinteraksi dengan pengguna lain dan berbagi informasi melalui komunitas dan jejaring sosial. Banyaknya unggahan dari milyaran pengguna media sosial menjadi sumber data untuk mengekstrak dan membuat informasi baru. Penelitian dimulai dengan membagikan formulir kesediaan dan kuesioner untuk mendapatkan persetujuan dari responden yang menggunakan bahasa Indonesia di tweet mereka untuk berpartisipasi dalam penelitian ini. *Agglomerative Hierarchical Clustering* yang dipilih untuk memperkaya metode prediksi kepribadian seseorang berdasarkan konten di media sosial. Model pada penelitian ini memiliki akurasi 20.1% dengan rata-rata *silhouette score* -0.23. Keunikan kata yang tinggi dari setiap tweet yang diproses menjadi tantangan bagi model ini untuk menghasilkan performa yang optimal. Model ini dapat menangani data dalam jumlah besar dalam waktu singkat tetapi belum memberikan performa yang lebih optimal dibandingkan kasus serupa yang diselesaikan dengan *supervised learning*.

Kata kunci: media sosial, kepribadian, prediksi, metode, tweet