

Abstrak

Twitter merupakan layanan mikroblog yang memungkinkan pengguna dapat mengirim dan membaca *tweet* atau pesan dengan batasan hanya 280 karakter per *tweet*. Hal ini menyebabkan *tweet* yang ditulis cukup singkat, tidak selalu menggunakan tata Bahasa yang benar, dan sering menggunakan Bahasa gaul. Seringkali, percobaan klasifikasi topik dalam *tweet* mengalami akurasi rendah dikarenakan faktor-faktor tersebut. Maka dari itu, penelitian ini mengimplementasikan ekspansi fitur untuk mengurangi ketidakcocokan kata dan mengurangi fitur yang bernilai 0 menjadi nilai dari *similarity* katanya jika *similarity* katanya muncul pada *tweet* tersebut. Proses ekspansi fitur tersebut dapat membuat vektor representasi *tweet* yang ukuran dimensinya cukup besar dan *sparse* bisa membuat mesin mendapat semantik informasinya dan menghasilkan akurasi yang bagus. Metode ekspansi fitur ini mencari *similarity* dari katanya menggunakan *fastText*. Hasil penelitian menunjukkan bahwa sistem klasifikasi topik menggunakan metode *Gradient Boosted Decision Tree* dan dengan ekspansi fitur memiliki akurasi tertinggi yaitu 91,39%.