

ABSTRAK

Twitter merupakan mikroblog yang sedang diminati oleh banyak orang di berbagai negara. Para penggunanya dapat memposting segala hal mengenai apa yang sedang terjadi di sekitar mereka. Baik itu *breaking news*, berita gosip selebriti, dan lain sebagainya.

Setiap harinya, tidak kurang dari 500 juta *tweet* dikirimkan oleh penggunanya dari seluruh penjuru dunia. Dari *tweet-tweet* tersebut, *Twitter* akan mendaftarkan topik mana saja yang sering dibicarakan secara *real-time* melalui fitur *Trending Topic*. Namun, pada kenyataannya tren topik tersebut mengacu pada banyak, dan seringnya suatu kata, atau frase dikicaukan oleh pengguna di berbagai lokasi mengenai suatu *event* besar yang sedang terjadi.

Penelitian ini bertujuan untuk mengklasifikasikan *tweet-tweet* yang dikirimkan dari kota Jakarta, Bandung, dan Makassar ke dalam beberapa topik, kemudian melihat tingkat kepopuleran topik tersebut di berbagai paruh waktu.

Penelitian ini diawali dengan pengumpulan data *tweet* dari *twitter.com*, kemudian dilakukan pra-proses. Dilanjutkan ke tahap pembentukan *classifier* memanfaatkan data latih. Tahap terakhir adalah klasifikasi *tweet* ke dalam parameter topik tertentu menggunakan metode Multinomial Naïve Bayes Classifier. Penelitian ini menghasilkan rata-rata *f1-score* yang cukup tinggi, yaitu 76,81% jika dibandingkan dengan *classifier* lain seperti Random Forest, SVM, dan Nearest Neighbor (75,18%, 68,82%, dan 50,22%).

Kata kunci: Klasifikasi topik, Twitter, Jakarta, Bandung, Makassar, Multinomial Naïve Bayes Classifier