

ABSTRAK

Pandemi Covid-19 melanda seluruh dunia tanpa terkecuali Indonesia, dengan angka kematian yang tinggi maka hal ini menimbulkan banyak tanggapan dari masyarakat yang mereka sampaikan melalui sosial media selama pandemi. *Twitter* merupakan salah satu sosial media yang digunakan untuk menyampaikan tanggapan berupa *tweet*. Namun dengan banyaknya tanggapan, diperlukan waktu lebih untuk menentukan apa saja informasi penting didalamnya. Untuk itu pada penelitian ini dikembangkan sistem yang mengidentifikasi topik pada *tweet*, topik tersebut berupa tanggapan masyarakat dari sosial media *twitter*. Identifikasi topik dalam *text mining* disebut *Topic Modelling*, pada skripsi ini tahapan dari *Topic Modeling* terdiri dari *Crawling Data*, *Text - Preprocessing* pada bagian *Cleansing* terdiri dari (*Case Folding*, *Tokenizing*, *Stopword Removal*, dan *Stemming*), pada bagian pembobotan kata menggunakan *Term Frequency-Inverse Document Frequency* (TF-IDF). *Text - Preprocessing* pada bagian seleksi fitur menggunakan *Pearson Correlation* dengan *threshold* lebih dari 0.6 dengan tujuan untuk mengurangi jumlah term yang memiliki nilai korelasi dibawah 0.6. *Topic Modeling* dengan algoritma *Latent Semantic Analysis* dengan *output* berupa *keyword* yang dihasilkan kemudian diberi *label* oleh pakar, evaluasi topik menggunakan *Topic Coherence* semakin mendekati 1, maka semakin bagus kualitas topik yang dihasilkan. Dari skenario uji coba yang dilakukan terhadap nilai *threshold* 0.6 – 0.9 dengan data hasil *crawling* “vaksinasi” dan “kebijakan vaksinasi” ditemukan hanya satu topik dengan label “Politik” dan *Coherence Score* yang berbeda yaitu, untuk nilai *threshold* 0.6 dengan *Coherence score* = 0.0004955403009978006, untuk nilai *threshold* 0.7 dengan *Coherence score* = 0, sedangkan nilai *threshold* 0.8 dan 0.9 menghasilkan nilai *Coherence score* yang sama yaitu 0.

Kata Kunci : *Topic Modelling, Latent Semantic Analysis, Twitter*